	Are Minimised In Response To Reactive Oxygen	Activities Of Pathways Not Maintained Under Reactive Oxygen Programmed Cell Death	- Many Proteins In Cells Undergoing Cell Death Or In Damaged Cells

Further, promoters of reactive oxygen responsive genes, as described in the Reference tables, for example, are useful to modulate transcription that is induced by reactive oxygen or any of the following phenotypes or biological activities below.

V.G. SALICYLIC ACID RESPONSIVE GENES, GENE COMPONENTS AND PRODUCTS

Plant defense responses can be divided into two groups: constitutive and induced. Salicylic acid (SA) is a signaling molecule necessary for activation of the plant induced defense system known as systemic acquired resistance or SAR. This response, which is triggered by prior exposure to avirulent pathogens, is long lasting and provides protection against a broad spectrum of pathogens. Another induced defense system is the hypersensitive response (HR). HR is far more rapid, occurs at the sites of pathogen (avirulent pathogens) entry and precedes SAR. SA is also the key signaling molecule for this defense pathway.

Changes in SA concentration in the surrounding environment or within a plant results in modulation of many genes and gene products. Examples of such SA responsive genes and gene products are shown in the Reference, Sequence, Protein Group, Protein Group Matrix, MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff and MA_clust tables. These genes and/or products are responsible for effects on traits such as plant vigor and seed yield. They were discovered and characterized from a much larger set by experiments designed to find genes whose mRNA products changed in response to SA treatment.

While SA responsive polynucleotides and gene products can act alone, combinations of these polynucleotides also affect growth and development. Useful combinations include different SA responsive polynucleotides and/or gene products that have similar transcription profiles or similar biological activities, and members of the same or similar biochemical pathways. In addition, the combination of SA responsive polynucleotides and/or gene product with another environmentally responsive polynucleotide is also useful because of the interactions that exist between hormone regulated pathways, stress and pathogen induced pathways, nutritional pathways and development. Here, in addition to polynucleotides having similar transcription profiles and/or biological activities, useful combinations include polynucleotides that may have different transcription profiles but which participate in common and overlapping pathways.

Such SA responsive genes and gene products can function to either increase or dampen the above phenotypes or activities either in response to changes in SA concentration or in the absence of SA fluctuations. The MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or

AFLP_diff Table(s) reports the transcript levels of the experiment (see EXPT ID: 108586, 108587, 108515, 108552, 108471, 108472, 108469, 108470, 107953, 107960, 108443, 108440, 108441, 108475, 108476). For transcripts that had higher levels in the samples than the control, a "+" is shown. A "-" is shown for when transcript levels were reduced in root tips as compared to the control. For more experimental detail see the Example section below.

SA genes are those sequences that showed differential expression as compared to controls, namely those sequences identified in the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff tables with a "+" or "-" indication.

SA Genes Identified By Cluster Analyses Of Differential Expression

SA Genes Identified By Correlation To Genes That Are Differentially Expressed

As described above, the transcription profiles of genes that act together are well correlated. Applicants not only have identified the genes that are differentially expressed in the microarray experiments, but also have identified the genes that act in concert with them. The MA_clust table indicates groups of genes that have well correlated transcription profiles and therefore participate in the same pathway or network.

A pathway or network of SA genes is any group in the MA_clust that comprises a cDNA ID that also appears in Expt ID 108586, 108587, 108515, 108552, 108471, 108472, 108469, 108470, 107953, 107960, 108443, 108440, 108441, 108475, 108476 of the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table(s).

SA Genes Identified By Correlation To Genes That Cause Physiological Consequences

Additionally, the differential expression data and the phenotypic observations can be merged to identify pathways or networks of SA genes. A group in the MA_clust is considered a SA pathway or network if the group comprises a cDNA ID that also appears in Knock-in or Knock-out tables that causes one or more of the phenotypes described in section above.

SA Genes Identified By Amino Acid Sequence Similarity

SA genes from other plant species typically encode polypeptides that share amino acid similarity to the sequences encoded by corn and Arabidopsis SA genes. Groups of SA genes are identified in the Protein Group table. In this table, any protein group that comprises a peptide ID

that corresponds to a cDNA ID member of a SA pathway or network is a group of proteins that also exhibits SA functions/utilities.

Further, promoters of SA responsive genes, as described in the Reference tables, for example, are useful to modulate transcription that is induced by SA or any of the following phenotypes or biological activities below.

V.G.1. USE OF SALICYLIC ACID-RESPONSIVE GENES TO MODULATE PHENOTYPES

SA responsive genes and gene products are useful to or modulate one or more of the following phenotypes:

- Pathogen Tolerance and/or Resistance
- Avr/R locus Interactions
- Non-Host Interactions
- HR
- SAR, e.g., SA Responsive Genes And/Or Products In Conjuction With Any Of The Organisms Listed Below
- Resistance To Bacteria e.g. to Erwinia stewartii, Pseudomonas syringae, Pseudomonas tabaci, Stuart's wilt, etc.
- Resistance To Fungi e.g. to Downy Mildews Such As Scleropthora macrospora, Sclerophthora rayissiae, Sclerospora graminicola, Peronosclerospora sorghi, Peronosclerospora philippinensis, Peronosclerospora sacchari, Peronosclerospora maydis; Rusts Such As Puccinia sorphi, Puccinia polysora, Physopella zeae, etc.; And To Other Fungal Diseases e.g. Cercospora zeae-maydis, Colletotrichum graminicola, Fusarium monoliforme, Exserohilum turcicum, Bipolaris maydis, Phytophthora parasitica, Peronospora tabacina, Septoria, etc.;
- Resistance To Viruses Or Viroids e.g., To Tobacco Or Cucumber Mosaic Virus, Ringspot Virus, Necrosis Virus, Pelargonium Leaf Curl Virus, Red Clover Mottle Virus, Tomato Bushy Stunt Virus, And Like Viruses;
- Resistance To Insecs, Such As To Aphids e.g. Myzus persicae; to Beetles And Beetle Larvae; to lepidoptera larvae e.g. Heliothus etc.

- Resistance to Nematodes, e.g. Meloidogyne incognita etc
- Local Resistance In Primary (Infected) Or Secondary (Uninfected) Leaves
- Stress Tolerance
- Winter Survival
- Cold Tolerance
- Salt Tolerance
- Heavy Metal Tolerance, Such As Cadmium
- Tolerance To Physical Wounding;
- Increased Organelle Tolerance To Redox Stress, Such As In Mitochondria, And Chloroplasts
- Cell Death
- Programmed Cell Death, Including Death Of Diseased Tissue And During Senescence
- Fruit Drop
- Biomass
- Fresh And Dry Weight During Any Time In Plant Life, Such As Maturation
- Number Of Flowers, Seeds, Branches, And/Or Leaves
- Seed Yield, Including Number, Size, Weight, And/Or Harvest Index
- Fruit Yield, Including Number, Size, Weight, And/Or Harvest Index
- Plant Development
- Time To Fruit Maturity
- Cell Wall Strengthening And Reinforcement
- Plant Product Quality
    - Paper Making Quality
- Food Additives
- Treatment Of Indications Modulated By Free Radicals
- Cancer

To regulate any of the desired phenotype(s) above, activities of one or more of the SA responsive genes or gene products can be modulated and the plants tested by screening for the desired trait. Specifically, the gene, mRNA levels, or protein levels can be altered in a plant utilizing the procedures described herein and the phenotypes can be assayed. As an example, a

plant can be transformed according to Bechtold and Pelletier (1998, Methods. Mol. Biol. 82:259-266) and/or screened for variants as in Winkler et al. (1998) Plant Physiol 118: 743-50 and visually inspected for the desired phenotype or metabolically and/or functionally assayed according to Zhao et al. (1998, Plant Cell 10:359-70) and Alvarez et al. (1998, Cell 92: 733-84).

V.G.2. USE OF SALICYLIC ACID-RESPONSIVE GENES TO MODULATE BIOCHEMICAL ACTIVITIES

The activities of one or more of the SA responsive genes can be modulated to change biochemical or metabolic activities and/or pathways such as those noted below. Such biological activities can be measured according to the citations included in the Table below:

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATION INCLUDING ASSAYS
Protection From Microbial Pathogens	Systemic Acquired Resistance (SAR) - Phytoalexin Biosynthesis - PR Protein Biosynthesis Local Resistance Wound Response	Alvarez et al. (1998) Cell 92: 733-84 Lapwood et al. (1984) Plant Pathol. 33: 13-20 Davis et al. (1993) Phytochemistry 32: 607-11 Yahraus et al. (1995) Plant Physiol. 109: 1259-66
Cell Signaling	- Modulation Of Reactive Oxygen Signaling - Modulation Of No Signaling	-Alvarez et al. (1998) Cell 92: 773-784 Delledonne et al. (1998) Nature 394: 585-588
Growth And Development	- Lignification	Redman et al. (1999) Plant Physiol. 119: 795-804

Other biological activities that can be modulated by the SA responsive genes and gene products are listed in the Reference tables. Assays for detecting such biological activities are described in the Protein Domain table.

Salicylic acid responsive genes are characteristically differentially transcribed in response to fluctuating SA levels or concentrations, whether internal or external to an organism or cell.

The MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table reports the changes in transcript levels of various SA responsive genes in entire seedlings at 1 and 6 hours after the seedling was sprayed with a Hoagland's solution enriched with SA as compared to seedlings sprayed with Hoagland's solution only.

The data from this time course can be used to identify a number of types of SA responsive genes and gene products, including "early responders" and "delayed responders." Profiles of these different SA responsive genes are shown in the Table below together with examples of the kinds of associated biological activities.

GENE EXPRESSION LEVELS	FUNCTIONAL CATEGORY OF GENE	PHYSIOLOGICAL CONSEQUENCES	EXAMPLES OF BIOCHEMICAL ACTIVITIES OF GENE PRODUCTS
Upregulated Genes (Level At 1h ≅6h) Or (Level At 1h > 6h)	Early Responders To SA	- SA Perception - SA Uptake - Modulation Of SA Response Transduction Pathways	-Transcription Factors -Transporters, Kinases, Phosphatases, G-Proteins, LRR, DNA Remodelling Proteins
Upregulated Genes (Level At 1h < 6h)	Delayed Responders To SA	- Specific Defensegene Transcription Initiation (E.G. Pr Genes, Pal	-Proteases, PRProteins, Cellulases, Chitinases, Cutinases, Other Degrading Enzymes, Pal, Proteins Of Defense Pathways, Cell Wall Proteins Epoxide Hydrolases, Methyl Transferases
Downregulated	- Early	- Negative Regulation	Transcription factors,

GENE EXPRESSION LEVELS	FUNCTIONAL CATEGORY OF GENE	PHYSIOLOGICAL CONSEQUENCES	EXAMPLES OF BIOCHEMICAL ACTIVITIES OF GENE PRODUCTS
(Level At 1h $\cong$ 6h) Or (Level At 6h >1h)	Responder Repressors To SA - Genes With Discontinued Expression Or UnsTable mRNA In The Presence Of SA	Of SA Inducible Pathways Released	kinases, phosphatases, G-proteins, LRR, transporters, calcium binding proteins, chromatin remodelling protein
Down-Regulated Transcripts (Level At 1h > 6h)	- Delayed Responders To SA Metabolism - Genes With Discontinued Expression Or UnsTable mRNA In The Presence Of SA	Negative Regulation Of SA Inducible Pathways Released	Transcription Factors, Kinases, Phosphatases, G-Proteins, LRR, Transporters, Calcium Binding Proteins, Chromatin Remodelling Protein

Further, any desired sequence can be transcribed in similar temporal, tissue, or enviromentally specific patterns as the SA responsive genes when the desired sequence is operably linked to a promoter of a SA responsive gene.

V.H. NITRIC OXIDE RESPONSIVE GENES, GENE COMPONENTS AND PRODUCTS

The rate-limiting element in plant growth and yield is often its ability to tolerate suboptimal or stress conditions, including pathogen attack conditions, wounding and the presence of various other factors. To combat such conditions, plant cells deploy a battery of inducible defense responses, including synergistic interactions between nitric oxide (NO), reactive oxygen intermediates (ROS), and salicylic acid (SA). NO has been shown to play a critical role in the activation of innate immune and inflammatory responses in animals. At least part of this mammalian signaling pathway is present in plants, where NO is known to potentiate the hypersensitive response (HR). In addition, NO is a stimulator molecule in plant photomorphogenesis.

Changes in nitric oxide concentration in the internal or surrounding environment, or in contact with a plant, results in modulation of many genes and gene products. Examples of such nitric oxide responsive genes and gene products are shown in the Reference and Sequence Tables. These genes and/or products are responsible for effects on traits such as plant vigor and seed yield. They were discovered and characterized from a much larger set by experiments designed to find genes whose mRNA products changed in response to nitric oxide treatment.

While nitric oxide responsive polynucleotides and gene products can act alone, combinations of these polynucleotides also affect growth and development. Useful combinations include different nitric oxide responsive polynucleotides and/or gene products that have similar transcription profiles or similar biological activities, and members of the same or similar biochemical pathways. Whole pathways or segments of pathways are controlled by transcription factor proteins and proteins controlling the activity of signal transduction pathways. Therefore, manipulation of the levels of such proteins is especially useful for altering phenotypes and biochemical activities of plants. In addition, the combination of a nitric oxide responsive polynucleotide and/or gene product with other environmentally responsive polynucleotides is also useful because of the interactions that exist between hormone regulated pathways, stress pathways, pathogen stimulated pathways, nutritional pathways and development. Here, in addition to polynucleotides having similar transcription profiles and/or biological activities, useful combinations include polynucleotides that may have different transcription profiles but which participate in common or overlapping pathways. The MA_diff and/or AFLP_diff MA_diff

and/or AFLP_diff and/or AFLP_diff Table(s) reports the transcript levels of the experiment (see EXPT ID: 108584, 108585, 108526, 108527, 108559). For transcripts that had higher levels in the samples than the control, a "+" is shown. A "-" is shown for when transcript levels were reduced in root tips as compared to the control. For more experimental detail see the Example section below.

NO genes are those sequences that showed differential expression as compared to controls, namely those sequences identified in the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff tables with a "+" or "-" indication

NO Genes Identified By Cluster Analyses Of Differential Expression

NO Genes Identified By Correlation To Genes That Are Differentially Expressed

As described above, the transcription profiles of genes that act together are well correlated. Applicants not only have identified the genes that are differentially expressed in the microarray experiments, but also have identified the genes that act in concert with them. The MA_clust table indicates groups of genes that have well correlated transcription profiles and therefore participate in the same pathway or network.

A pathway or network of NO genes is any group in the MA_clust that comprises a cDNA ID that also appears in Expt ID 108584, 108585, 108526, 108527, 108559 of the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table(s).

NO Genes Identified By Correlation To Genes That Cause Physiological Consequences

Additionally, the differential expression data and the phenotypic observations can be merged to identify pathways or networks of NO genes. A group in the MA_clust is considered a NO pathway or network if the group comprises a cDNA ID that also appears in Knock-in or Knock-out tables that causes one or more of the phenotypes described in section above.

NO Genes Identified By Amino Acid Sequence Similarity

NO genes from other plant species typically encode polypeptides that share amino acid similarity to the sequences encoded by corn and Arabidopsis NO genes. Groups of NO

genes are identified in the Protein Group table. In this table, any protein group that comprises a peptide ID that corresponds to a cDNA ID member of a NO pathway or network is a group of proteins that also exhibits NO functions/utilities.

Such nitric oxide responsive genes and gene products can function either to increase or dampen the above phenotypes or activities either in response to changes in nitric oxide concentration or in the absence of nitric oxide fluctuations. Further, promoters of nitric oxide responsive genes, as described in the Reference tables, for example, are useful to modulate transcription that is induced by nitric oxide or any of the following phenotypes or biological activities below.

V.H.1. USE OF NITRIC OXIDE-RESPONSIVE GENES TO MODULATE PHENOTYPES :

Nitric oxide responsive genes and gene products are useful to or modulate one or more of the following phenotypes:

- Stress Responses
- Mediation of response to stresses
- Disease resistance
- Growth
- Roots
- Stems
- Leaves
- Cells
- Promotes leaf cell elongation
- Biomass
- Fresh and Dry Weight during any time in plant life, such as at maturation;
- Size and/or Weight
- Flowers
- Seeds
- Branches

- Leaves
- Roots
- Development
- Seed Development
- Dormancy
- Control rate and timing of germination
- Prolongs seed storage and viability
- Senescence

Further, any desired sequence can be transcribed in similar temporal, tissue, or enviromentally specific patterns as the nitric responsive genes when the desired sequence is operably linked to a promoter of a nitric responsive gene.

To regulate any of the desired phenotype(s) above, activities of one or more of the nitric oxide responsive genes or gene products can be modulated and the plants tested by screening for the desired trait. Specifically, the gene, mRNA levels, or protein levels can be altered in a plant utilizing the procedures described herein and the phenotypes can be assayed. As an example, a plant can be transformed according to Bechtold and Pelletier (1998) Methods. Mol. Biol. 82: 259-266 and/or screened for variants as described in Winkler et al. (1998) Plant Physiol. 118: 743-50 and visually inspected for the desired phenotype. Alternatively, plants can be metabolically and/or functionally assayed according to Beligni and Lamattina (2000) Planta 210: 215-21), Lapwood et al (1984) Plant Pathol 33: 13-20, and/or Brown and Botstein (1999) Nature Genet. 21: 33-37.

V.H.2. USE OF NITRIC OXIDE-RESPONSIVE GENES TO MODULATE BIOCHEMICAL ACTIVITIES:

The activities of one or more of the nitric oxide responsive genes can be modulated to change biochemical or metabolic activities and/or pathways such as those noted below. Such biological activities can be measured according to the citations included in the Table below:

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
Stress Response	-Programmed Cell Death	Levine et al (1996) Curr. Biol 6: 427-37 Sellins and Cohen (1991) Radiat. Res. 126: 88-95
	-Reactive Oxygen based Defence Pathways	Kumar and Klessig (2000) Mol. Plant Microbe Interact. 13:347-351
Disease Resistance	-Microbial Pathogen resistance pathways	Lapwood et al (1984) Plant Pathol 33: 13-20 Kumar and Klessig (2000) Mol. Plant microbe interact.13: 347-351 Klessig et.al.(2000) Proc. Nat. Acad. Sci USA 97: 8849-8855 Delledonna et al(1998) Nature 394: 585-588
	-Programmed Cell Death	Levine et al (1996) Curr. Biol 6: 427-437 Sellins and Cohen (1991) Radiat. Res. 126: 88-95
	-Cellular Protectant Gene expression	Brown and Botstein (1999) Nat Genet 21: 33-37
	- Phytoalexin Biosynthesis	Davis et al. (1993) Phytochemistry 32: 607-

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
		611

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
Signal Transduction	Regulation of hydrogen peroxide signaling	Wu et al. (1995) Plant Cell 7, 1357-1368
Reorientation of nitrogen metabolism	Induction of ribosomal proteins, asparagine synthesis, proteases, Rnases	This study. Standard assays for detection of changes
Reorientation of sugar and energy metabolism	Induction of sugar transporters, ATPases, glycohydrolases, and glycolytic enzymes, for example	This study. Standard assays for detection of changes

Other biological activities that can be modulated by the NO responsive genes and gene products are listed in the Reference Tables. Assays for detecting such biological activities are described in the Protein Domain table.

NO responsive genes are characteristically differentially transcribed in response to fluctuating NO levels or concentrations, whether internal or external to an organism or cell. The MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table(s) report(s) the changes in transcript levels of various NO responsive genes in aerial tissues at 1 and 6 hours after a plant was sprayed with a Silwett L-77 solution enriched with 5 mM sodium nitroprusside, which is an NO donor. These changes are in comparison with plants sprayed with Silwett L-77 solution only.

The data from this time course can be used to identify a number of types of NO responsive genes and gene products, including "early responders" and "delayed responders" Profiles of these different nitric oxide responsive genes are shown in the Table below together with examples of the kinds of associated biological activities.

GENE EXPRESSION LEVEL	FUNCTIONAL CATEGORY OF GENE	PHYSIOLOGICAL CONSEQUENCES	EXAMPLES OF BIOCHEMICAL ACTIVITY
Upregulated genes (level at 1 hour $\cong$ 6 hours) (level at 1 hour > 6 hours)	Early responder repressors to NO	- NO Perception - NO Uptake - Modulation of NO Response Transduction Pathways Specific Gene Transcription Initiation of Pathways to Optimize NO Response Pathways	-Transcription Factors -Transporters -Pathogen responsive proteins, salicylic and jasmonate pathway proteins -Proteins to provide defence against active oxygen e.g. glutathione transferase,ascorbate free radical reductase, ascorbate peroxidase, nitrilase, heat shock proteins -Proteins to reorient metabolism e.g.proteases, Rnases, proteasomes, asparagine synthetase, glycohydrolases, transporters -Proteins to inhibit transport of nitric oxide -Degradation enzymes

GENE EXPRESSION LEVEL	FUNCTIONAL CATEGORY OF GENE	PHYSIOLOGICAL CONSEQUENCES	EXAMPLES OF BIOCHEMICAL ACTIVITY
Upregulated transcripts (level at 1 hour < 6 hours)	Delayed NO responders	- Maintenance of metabolism in presence of High NO -Maintenace of disease defence pathways -Maintenance of pathways against reactive oxygen production Maintenance of different metabolic programs Selective cell death	- NO Metabolic Pathway enzymes -Pathogen responsive proteins, salicylic and jasmonate pathway proteins -Proteins to provide defence against active oxygen e.g. glutathione transferase,ascorbate free radical reductase, ascorbate peroxidase, nitrilase, heat shock proteins -Proteins to reorient and sustain metabolism e.g.proteases,Rnases,proteasomes,asparagine synthetase,glycohydrolases,transporters, -Proteins to inhibit transport of NO -Degradation enzymes
Down Regulated Transcripts (level at 1 hours ≅ 6	-Early responders of NO utilization pathways	Negative regulation of NO utilization pathways released	-Transcription factors -Kinases and phosphatases

GENE EXPRESSION LEVEL	FUNCTIONAL CATEGORY OF GENE	PHYSIOLOGICAL CONSEQUENCES	EXAMPLES OF BIOCHEMICAL ACTIVITY
hours) (level at 6 hours > 1 hour)	-Genes with discontinued expression or unsTable mRNA following nitric oxide uptake	Reorientation of metabolism	-Chromatin restructuring proteins - Transcription factors,metabolic enzymes, kinases and phosphatases, transporters, ribosomal proteins
		Programmed cell death	-Most proteins in cells undergoing cell death
Down Regulated Transcripts (level at 1 hour > 6 hours)	-Delayed responder repressors of NO stress metabolism	Negative regulation of NO utilization pathways released	Transcription factors -Kinases and phosphatases -Chromatin restructuring proteins
	-Genes with discontinued expression or unsTable mRNA following nitric oxide uptake	Reorientation of metabolism	-Transcription factors,metabolic enzymes, kinases and phosphatases, transporters, ribosomal proteins.
		Programmed cell death	-Most proteins in cells undergoing programmed cell death

Use of Promoters of NO Responsive Genes

Promoters of NO responsive genes are useful for transcription of any desired polynucleotide or plant or non-plant origin. Further, any desired sequence can be transcribed in a similar temporal, tissue, or environmentally specific patterns as the NO responsive genes where the desired sequence is operably linked to a promoter of a NO responsive gene. The protein product of such a polynucleotide is usually synthesized in the same cells, in response to the same stimuli as the protein product of the gene from which the promoter was derived. Such promoter are also useful to produce antisense mRNAs to down-regulate the product of proteins, or to produce sense mRNAs to down-regulate mRNAs via sense suppression.

V.I. OSMOTIC STRESS RESPONSIVE GENES, GENE COMPONENTS AND PRODUCTS

The ability to endure and recover from osmotic and salt related stress is a major determinant of the geographical distribution and productivity of agricultural crops. Osmotic stress is a major component of stress imposed by saline soil and water deficit. Decreases in yield and crop failure frequently occur as a result of aberrant or transient environmental stress conditions even in areas considered suitable for the cultivation of a given species or cultivar. Only modest increases in the osmotic and salt tolerance of a crop species would have a dramatic impact on agricultural productivity. The development of genotypes with increased osmotic tolerance would provide a more reliable means to minimize crop losses and diminish the use of energy-costly practices to modify the soil environment.

Changes in the osmotic concentration of the surrounding environment or within a plant results in modulation of many genes and gene products. Examples of such osmotic stress responsive genes and gene products, including salt responsive genes, are shown in the Reference, Sequence, Protein Group, Protein Group Matrix, MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff and MA_clust tables. These genes and/or products are responsible for effects on traits such as plant vigor and seed yield.

While osmotic and/or salt stress responsive polynucleotides and gene products can act alone, combinations of these polynucleotides also affect growth and development. Useful combinations include different osmotic stress responsive polynucleotides and/or gene products that have similar transcription profiles or similar biological activities, and members of the same or similar biochemical pathways. In addition, the combination of an osmotic stress responsive polynucleotide and/or gene product with another environmentally responsive polynucleotide is also useful because of the interactions that exist between hormone regulated pathways, stress pathways, nutritional pathways and development. Here, in addition to polynucleotides having similar transcription profiles and/or biological activities, useful combinations include polynucleotides that may have different transcription profiles but which participate in a common pathway.

Such osmotic and/or salt stress responsive genes and gene products can function to either increase or dampen the above phenotypes or activities either in response to changes in osmotic concentration or in the absence of osmotic fluctuations. The MA_diff and/or AFLP_diff

MA_diff and/or AFLP_diff and/or AFLP_diff Table(s) reports the transcript levels of the experiment (see EXPT ID: 108570, 108571, 108541, 108542, 108553, 108539, 108540). For transcripts that had higher levels in the samples than the control, a "+" is shown. A "-" is shown for when transcript levels were reduced in root tips as compared to the control. For more experimental detail see the Example section below.

Osmotic Stress genes are those sequences that showed differential expression as compared to controls, namely those sequences identified in the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff tables with a "+" or "-" indication

Osmotic Stress Genes Identified By Cluster Analyses Of Differential Expression

Osmotic Stress Genes Identified By Correlation To Genes That Are Differentially Expressed

As described above, the transcription profiles of genes that act together are well correlated. Applicants not only have identified the genes that are differentially expressed in the microarray experiments, but also have identified the genes that act in concert with them. The MA_clust table indicates groups of genes that have well correlated transcription profiles and therefore participate in the same pathway or network.

A pathway or network of Osmotic Stress genes is any group in the MA_clust that comprises a cDNA ID that also appears in Expt ID 108570, 108571, 108541, 108542, 108553, 108539, 108540 of the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table(s).

Osmotic Stress Genes Identified By Correlation To Genes That Cause Physiological Consequences

Additionally, the differential expression data and the phenotypic observations can be merged to identify pathways or networks of Osmotic Stress genes. A group in the MA_clust is considered a Osmotic Stress pathway or network if the group comprises a cDNA ID that also appears in Knock-in or Knock-out tables that causes one or more of the phenotypes described in section above.

Osmotic Stress Genes Identified By Amino Acid Sequence Similarity

Osmotic Stress genes from other plant species typically encode polypeptides that share amino acid similarity to the sequences encoded by corn and Arabidopsis Osmotic Stress genes. Groups of Osmotic Stress genes are identified in the Protein Group table. In this table, any protein group that comprises a peptide ID that corresponds to a cDNA ID member of a Osmotic Stress pathway or network is a group of proteins that also exhibits Osmotic Stress functions/utilities.

Further, promoters of osmotic stress responsive genes, as described in the Reference tables, for example, are useful to modulate transcription that is induced by osmotic stress or any of the following phenotypes or biological activities below.

V.I.1. USE OF OSMOTIC STRESS RESPONSIVE GENES TO MODULATE PHENOTYPES

Osmotic stress responsive genes and gene products are useful to or modulate one or more of the following phenotypes:

- Growth
- Roots
- Stems
- Leaves
- Development
    - Cell Growth
        - DNA Synthesis and Cell Division
    - Seed Development
        - Desiccation tolerance
        - Dormancy
            - Control rate of Germination
            - Prolongs Seed Storage and Viability
    - Senescence
- Stress Responses
- Desiccation
- Drought

- Salt

To regulate any of the phenotype(s) above, activities of one or more of the osmotic stress responsive genes or gene products can be modulated and the plants tested by screening for the desired trait. Specifically, the gene, mRNA levels, or protein levels can be altered in a plant utilizing the procedures described herein and the phenotypes can be assayed. As an example, a plant can be transformed according to Bechtold and Pelletier (1998, Methods. Mol. Biol. 82:259-266) and/or screened for variants as in Winkler et al. (1998) Plant Physiol 118: 743-50 and visually inspected for the desired phenotype or metabolically and/or functionally assayed according to de Castro (1998, Phytochemistry 47: 689-694), Xu (1998, J Exp Bot 49: 573-582), Ausubel et al. (In: Current Protocols in Molecular Biology (1999) Volume 1, chapter 4, eds. Ausubel, Brent, Kingston, Moore, Seidman, Smith and Struhl, New York, NY) and De Castro et al. (2000, Plant Physiol 122: 327-36)

V.I.2. USE OF OSMOTIC STRESS RESPONSIVE GENES TO MODULATE BIOCHEMICAL ACTIVITIES

The activities of one or more of the osmotic stress responsive genes can be modulated to change biochemical or metabolic activities and/or pathways such as those noted below. Such biological activities can be measured according to the citations included in the Table below:

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
Cell Growth And Differentiation	Regulation Of Osmolyte Synthesis	Yoshu et al. (1995) The Plant Journal 7: 751-60
	Regulation Of Glycolate Pathway And Photoinhibition Of Photosystem II In Response To Stress	Streb et al. (1993) Physiologia Plantarum. 88:590-598

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
Gene Regulation	Transcriptional Regulation Of Osmotic Stress Induced Proteins Through DNA Binding Proteins	Current Protocols in Molecular Biology / edited by Frederick M. Ausubel .. [et al.]. New York : Published by Greene Pub. Associates and Wiley-Interscience : J. Wiley, c1987
	Transcriptional Regulation Of Osmotic Stress Induced Proteins Through Protein Phosphorylation And Dephosphorylation	Jonak (1996) Proceedings of the National Academy of Sciences of the United States of America, 93: 11274-11279; Monroy, A.et al., (1998) Analytical Biochemistry 265: 183-185;
	Regulation Of Osmotic Stress Induced Gene Protein Accumulation By Protein Protein Intereaction Between Osmotic Stress Regulated Genes And Protein Phosphatase 2C	McCright (1998) IN: Methods in Molecular Biology; Protein phosphatase protocols; Ludlow (1998) Humana Press Inc. ; Suite 808, 999 Riverview Drive, Totowa, New Jersey 07512, USA. :263-277.
	Transcriptional Regulation Of Heat Induced Genes Through	Luo and Dean (1999) Journal of the

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
	Chromatin Remodeling	National Cancer Institute 91: 1288-1294; Chromatin protocols (1999) edited by Peter B. Becker. Totowa, N.J. : Humana Press
	Activity Of Abcisic Acid Regulated DNA Binding Proteins	Gubler et al. (1999) Plant Journal 17: 1-9
	Accumulation Of RNA Binding Proteins That Regulate Osmotic Stress	Sato (1995) Nucleic Acids Research 23: 2161-2167.
Stress Response	Synthesis And Metabolism Of Osmoprotectants Such As Betaine, Proline And Trehalase	Minocha et al. (1999) Plant Physiol and Biochem 37: 597-603
	Regulation Of Sugar Transporters	Dejardin et al. (1999) Biochem J; 344 Pt 2:503-9
	Regulation Of Vacuolar Sodium/Proton Antiport Activity And The Detoxification Of Cations	Gaxiola et al. (1999) PNAS USA 96: 1480-1485
	Regulation Of Intracellular Na+	Espinoza-Ruiz et

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
	And Li+ Ion Concentrations	al. (1999) The Plant Journal 20: 529-539
	Regulation Of Universal Stress Protein Homologue Activity By Phosphorylation And Dephosphorylation.	Freestone et al. (1997) Journal of Molecular Biology, v. 274: 318-324
	Regulation/Maintenance Of Protein Stability During Thermal Stress	Walker (1996) Humana Press Inc. Suite 808, 999 Riverview Drive, Totowa, New Jersey 07512, USA
	Regulation Of Protein Degradation During Thermal Stress.	Vierstra (1996) Plant Molecular Biology,32:275-302. Vierstra and Callis (1999) Plant Molecular Biology, 41:435-442
Signal Transduction	Activation Of Stress Response Genes	Xinong et al. (1999) The Plant Journal 19: 569-578
	Salt Tolerance	Piao (1999) Plant Physiol 19: 1527-1534
	Calcium Mediated Stress Response	Subbaiah et al. (1994) Plant Physiology

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
		105:369-376 Kudla et al. (1999) PNAS USA 96: 4718-4723

Other biological activities that can be modulated by the osmotic stress responsive genes and gene products are listed in the Reference tables. Assays for detecting such biological activities are described in the Protein Domain table.

Osmotic stress responsive genes are characteristically differentially transcribed in response to fluctuating osmotic stress levels or concentrations, whether internal or external to an organism or cell. MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table reports the changes in transcript levels of various osmotic stress responsive genes in aerial tissues of plants at 1 and 6 hours after the plants were sprayed with Hoagland's solution containing 20% PEG as compared to aerial tissues from plants sprayed with Hoagland's solution only.

The data from this time course can be used to identify a number of types of osmotic stress responsive genes and gene products, including "early responding," "sustained osmotic stress responders," "repressors of osmotic stress pathways" and "osmotic stress responders." Profiles of these different osmotic stress responsive genes are shown in the Table below together with examples of the kinds of associated biological activities.

GENE EXPRESSION LEVELS	FUNCTIONAL CATEGORY OF GENES	PHYSIOLOGICAL CONSEQUENCES	EXAMPLES OF BIOCHEMICAL ACTIVITIES OF GENE PRODUCTS
Up Regulated Transcripts (Level At 1 Hour ≅ 6 Hours)	• Early Responders To Osmotic Stress	• Osmotic Stress Perception • Osmolyte Uptake • Modulation Of	• Transcription Factors • Transcription Coactivators

(Level At 1 Hour > 6 Hours)	• Universal Stress Response Genes • Osmotic Stress Responders • Abscisic Acid Biosynthesis And Perception	Osmotic Stress Response Signal Transduction Pathways • Specific Gene Transcription Initiation • Specific Gene Transcription Repression • Translation Activation • Translation Repression • Repression Of "Normal State" Pathways To Optimize Osmotic Stress Response • Activation Of Stress Signaling Pathways • Up Regulation Of Abscisic Acid Biosynthesis Pathway Protein Accumulation And Activity • Scavenging Reactive Oxygen Species • Modification Of Cell Wall Composition • Up-Regulation Of Universal Stress Response Protein	• Membrane Transporters • Proline Biosynthesis • Selective Inhibition Of Osmolyte Transport • Protein Ubiquitination • Protein Degradation • Rna Binding Proteins • Modification Of Protein Activity By Phosphatases, Kinases • Synthesis And Or Activation Of Oxide Hydrolases, Suoeroxidedismutase, Iron Ascorbate Peroxidase • Activation Of Signaling Pathway By Calcium Binding Proteins, • Modification Of Protein Activity By Protein-Protein Interaction

		Accumulation	• Change In Chromatin Structure And/Or Localized Dna Topology • Modification Of Pre-Existing Translation Factors By Phosphorylation (Kinases) Or Dephosphorylation (Phosphatases) • Synthesis Of New Translation Factors • Abscisic Acid Biosynthesis
Up Regulated Transcripts (Level At 1 Hr < 6 Hr)	• Sustained Osmotic Stress Responders • Repressor Of Osmotic Stress Pathways • Abscisic Acid Perception, Biosynthesis And Regulation	• Osmolyte Adjustment And Adaptation • Photosynthetic Activity Modification • Activation Of "Normal State" Biosynthesis Genes • Negative Regulation Of Osmotic Stress Pathways • Negative Regulation Of Abscisic Acid Biosynthesis • Acivation Of Abscisic Acid Degradation	• Osmotic Stress Metabolic Pathways • Sugar Biosynthetic Pathways • Sugar Transporters • Transcription Factors • Transcription Coactivators • Membrane Transporters • Abscisic Acid Biosynthesis

		Pathway • Cell Wall Composition Modification	
Down-Regulated Transcripts (Level At 1 Hr ≈ 6 Hr) (Level At 6 Hr > 1 Hr)	Early Responder Repressors Of "Normal" State Of Metabolism Negative Regulators Of Abscisic Acid Biosynthesis And Perception. Positive Regulators Of "Normal State" Metabolic Pathways.	• Metabolic Repression • Specific Gene Transcription Initiation • Specific Gene Transcription Repression • Translation Activation • Translation Repression • Abscisic Acid Degradation • Protein Degradation	• Transcription Factors • Transcription Coactivators • Protein Degradation • Rna Binding Proteins • Modification Of Protein Activity By Phosphatases, Kinases • Activation Of Signaling Pathway By Calcium Binding Proteins, • Modification Of Protein Activity By Protein-Protein Interaction • Change In Chromatin Structure And/Or Localized Dna Topology • Modification Of Pre-Existing

			Translation Factors By Phosphorylation (Kinases) Or Dephosphorylation (Phosphatases) • Synthesis Of New Translation Factors
Down-Regulated Transcripts (Level At 1 Hr > 6 Hr)	Repressors Of "Normal" State Of Metabolism Genes With Discontinued Expression Or UnsTable mRNA In Presence Of Osmotic Stress Repressor Of Osmotic Stress Pathways Repressors Of Abscisic Acid Biosynthesis, Perception And Regulation	• Osmotic Stress Adaptation • Negative Regulation Of Abscisic Acid Biosynthesis • Negative Regulation Of Osmotic Stress Response Pathways Genes • Osmolyte Synthesis And Osmolyte Cellular Partitioning Readjustment • Activation Of "Normal State" Metabolic Pathways	• Transcription Factors • Transcription Coactivators • Protein Degradation • Rna Binding Proteins • Modification Of Protein Activity By Phosphatases, Kinases • Activation Of Signaling Pathway By Calcium Binding Proteins, • Modification Of Protein Activity By Protein-Protein Interaction • Change In Chromatin Structure And/Or Localized Dna

			Topology • Modification Of Pre-Existing Translation Factors By Phosphorylation (Kinases) Or Dephosphorylation (Phosphatases) • Synthesis Of New ○ Translation Factors • Sugar Biosynthetic Pathways • Sugar Transporters

Further, any desired sequence can be transcribed in similar temporal, tissue, or enviromentally specific patterns as the osmotic stress responsive genes when the desired sequence is operably linked to a promoter of an osmotic stress responsive gene.

V.J. ALUMINUM RESPONSIVE GENES, GENE COMPONENTS AND PRODUCTS

Aluminum is toxic to plants in soluble form (Al^{3+}). Plants grown under aluminum stress have inhibited root growth and function due to reduced cell elongation, inhibited cell division and metabolic interference. As an example, protein inactivation frequently results from displacement of the Mg2+ cofactor with aluminum. These types of consequences result in poor nutrient and water uptake. In addition, because stress perception and response occur in the root apex, aluminum exposure leads to the release of organic acids, such as citrate, from the root as the plant attempts to prevent aluminum uptake.

The ability to endure soluble aluminum is a major determinant of the geographical distribution and productivity of agricultural crops. Decreases in yield and crop failure frequently occur as a result of aberrant, hot conditions even in areas considered suiTable for the cultivation of a given species or cultivar. Only modest increases in the aluminum tolerance of crop species would have a dramatic impact on agricultural productivity. The development of genotypes with increased aluminum tolerance would provide a more reliable means to minimize crop losses and diminish the use of costly practices to modify the environment.

Microarray technology allows monitoring of gene expression levels for thousands of genes in a single experiment. This is achieved by simultaneously hybridizing two differentially labeled fluorescent cDNA pools to glass slides that contain spots of DNA (Schena et al. (1995) Science 270: 467-70). The Arabidopsis Functional Genomics Consortium (AFGC) has recently made public the results from such microarray experiments conducted with AFGC chips containing 10,000 non-redundant ESTs, selected from 37,000 randomly sequenced ESTs generated from mRNA of different tissues and developmental stages.

The sequences of the ESTs showing at least two-fold increases or decreases over the controls were identified, compared to the Ceres full-length cDNA and genomic sequence databanks, and identical Ceres clones identified. MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table reports the results of this analysis, indicating those Ceres clones which are up or down regulated over controls, thereby indicating the Ceres clones which are aluminum response responsive genes.

The MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff Table(s) reports the transcript levels of the experiment (see EXPT ID: Aluminum (relating to SMD 7304,

SMD 7305)). For transcripts that had higher levels in the samples than the control, a "+" is shown. A "-" is shown for when transcript levels were reduced in root tips as compared to the control. For more experimental detail see the Example section below.

Aluminum genes are those sequences that showed differential expression as compared to controls, namely those sequences identified in the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff tables with a "+" or "-" indication

Aluminum Genes Identified By Cluster Analyses Of Differential Expression

Aluminum Genes Identified By Correlation To Genes That Are Differentially Expressed

As described above, the transcription profiles of genes that act together are well correlated. Applicants not only have identified the genes that are differentially expressed in the microarray experiments, but also have identified the genes that act in concert with them. The MA_clust table indicates groups of genes that have well correlated transcription profiles and therefore participate in the same pathway or network.

A pathway or network of Aluminum genes is any group in the MA_clust that comprises a cDNA ID that also appears in Expt ID Aluminum (relating to SMD 7304, SMD 7305) of the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table(s).

Aluminum Genes Identified By Correlation To Genes That Cause Physiological Consequences

Additionally, the differential expression data and the phenotypic observations can be merged to identify pathways or networks of Aluminum genes. A group in the MA_clust is considered a Aluminum pathway or network if the group comprises a cDNA ID that also appears in Knock-in or Knock-out tables that causes one or more of the phenotypes described in section above.

Aluminum Genes Identified By Amino Acid Sequence Similarity

Aluminum genes from other plant species typically encode polypeptides that share amino acid similarity to the sequences encoded by corn and Arabidopsis Aluminum genes. Groups of Aluminum genes are identified in the Protein Group table. In this table, any protein

group that comprises a peptide ID that corresponds to a cDNA ID member of a Aluminum pathway or network is a group of proteins that also exhibits Aluminum functions/utilities.

V.J.1. USE OF ALUMINUM RESPONSE GENES TO MODULATE PHENOTYPES

Changes in aluminum concentrations in a plant's surrounding environment results in modulation of many genes and gene products. Examples of such aluminum response genes and gene products are shown in the Reference and Sequence Tables. These genes and/or products are responsible for effects on traits such as plant vigor and seed yield.

While aluminum responsive polynucleotides and gene products can act alone, combinations of these polynucleotides also affect growth and development. Useful combinations include different aluminum responsive polynucleotides and/or gene products that have similar transcription profiles or similar biological activities, and members of the same or similar biochemical pathways. In addition, the combination of a aluminum responsive polynucleotide and/or gene product with another environmentally responsive polynucleotide is also useful because of the interactions that exist between hormone regulated pathways, stress pathways, nutritional pathways and development. Here, in addition to polynucleotides having similar transcription profiles and/or biological activities, useful combinations include polynucleotides that may have different transcription profiles but which participate in a common pathway.

Such aluminum responsive genes and gene products can function to either increase or dampen the above phenotypes or activities either

- in response to changes in aluminum concentration or
- in the absence of aluminum fluctuations.

More specifically, aluminum responsive genes and gene products are useful to or modulate one or more of the following phenotypes:

- Growth
- Roots
    - Inhibition of root elongation
- Stems

- Leaves
- Whole Plant
- Development
    - Cell Growth
        - Elongation
        - Division
    - Mediates response to oxidative stress, calcium-mediated defense, antioxidant defense and pathogenesis

To produce the desired phenotype(s) above, one or more of the aluminum response genes or gene products can be tested by screening for the desired trait. Specifically, the gene, mRNA levels, or protein levels can be altered in a plant utilizing the procedures described herein and the phenotypes can be assayed. As an example, a plant can be transformed according to Bechtold and Pelletier (1998, Methods. Mol. Biol. 82:259-266) and visually inspected for the desired phenotype or metabolically and/or functionally assayed according to Li and Fleming (1999, FEBS Lett 461: 1-5), Delhaize et al. (1999, J Biol Chem 274: 7082-8), Sigimoto and Sakamoto (1997, Genes Genet Syst 72: 311-6), Esaki et al. (2000, Plant Physiol 122: 657-65), Leonard and Gerber (1988, Mutat Res 196: 247-57), Baisakhi et al. (2000, Mutat Res 465: 1-9), Ma (2000, Plant Cell Physiol 41: 383-90) and Koyama et al. (1999, Plant Cell 40: 482-8)

Alternatively, the activities of one or more of the aluminum responsive genes can be modulated to change biochemical or metabolic activities and/or pathways such as those noted below. Such biological activities can be measured according to the citations included in the Table below:

GENERAL CATEGORY	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	ASSAY
Cell Growth and Development	-Phospholipase D (PLD) activity	Toda et al. (1999) Biosci Biotechnol

	-Regulation of Phosphtidylserine Synthase (PSS) -Cell wall strengthening	Biochem 63: 210-212 Hamel et al. (1998) Planta 205: 531-38
Stress Response	-Regulation of oxidative stress -Regulation of antioxidant defense and DNA repair -Secretion of Organic Acids (e.g. maleate, citrate) from root apex -Ca2+mediated Defense Responses Against Low pH	Esaki et al. (2000) Plant Physiol 122: 657-655 Baisakhi et al. (2000) Mutat Res 465: 1-9 Koyama et al. (1999) Plant Cell 40: 482-8 Plieth et al. (1999) Plant J 18: 634-50
Signaling	-H+ transport -Auxin transport	Degenhardt et al. (1988) Plant Physil 117: 19-27 Rashotte et al. (2000) Plant Physiol 122: 481-90

Other biological activities that can be modulated by aluminum response genes and their products are listed in the REFERENCE Table. Assays for detecting such biological activities are described in the Protein Domain table.

TRANSCRIPT LEVELS	TYPE OF GENES	PHYSIOLOGICAL CONSEQUENCES	EXAMPLES OF BIOCHEMICAL ACTIVITY
Up regulated transcripts	responders to aluminum application	• Aluminum perception • Aluminum uptake and transport • Aluminum metabolism • Synthesis of secondary metabolites and/or proteins • Modulation of aluminum response transduction pathways • Specific gene transcription initiation	• Transporters • Metabolic enzymes • Change in cell membrane structure and potential • Kinases and phosphatases • Transcription activators • Change in chromatin structure and/or localized DNA topology •
Down-regulated transcripts	responder to aluminum repressors of aluminum state of metabolism Genes with	• Negative regulation of aluminum pathways • Changes in pathways and	• Transcription factors • Change in protein structure by phosphorylation (kinases) or dephosphorylation (phosphatases)

TRANSCRIPT LEVELS	TYPE OF GENES	PHYSIOLOGICAL CONSEQUENCES	EXAMPLES OF BIOCHEMICAL ACTIVITY
	discontinued expression or unsTable mRNA in presence of aluminum	processes operating in cells • Changes in other metabolisms than aluminum	• Change in chromatin structure and/or DNA topology • Stability of factors for protein synthesis and degradation • Metabolic enzymes

Use of Promoters of Aluminum Responsive Genes

Promoters of Aluminum responsive genes are useful for transcription of any desired polynucleotide or plant or non-plant origin. Further, any desired sequence can be transcribed in a similar temporal, tissue, or environmentally specific patterns as the Aluminum responsive genes where the desired sequence is operably linked to a promoter of a Aluminum responsive gene. The protein product of such a polynucleotide is usually synthesized in the same cells, in response to the same stimuli as the protein product of the gene from which the promoter was derived. Such promoter are also useful to produce antisense mRNAs to down-regulate the product of proteins, or to produce sense mRNAs to down-regulate mRNAs via sense suppression.

V.K. CADMIUM RESPONSIVE GENES, GENE COMPONENTS AND PRODUCTS

Cadmium (Cd) has both toxic and non-toxic effects on plants. Plants exposed to non-toxic concentrations of cadmium are blocked for viral disease due to the inhibition of systemic movement of the virus. Surprisingly, higher, toxic levels of Cd do not inhibit viral systemic movement, suggesting that cellular factors that interfere with the viral movement are triggered by non-toxic Cd concentrations but repressed in high Cd concentrations. Furthermore, exposure to non-toxic Cd levels appears to reverse posttranslational gene silencing, an inherent plant defense mechanism. Consequently, exploring the effects of Cd exposure has potential for advances in plant disease control in addition to soil bio-remediation and the improvement of plant performance in agriculture.

Changes in cadmium concentrations in a plant's surrounding environment results in modulation of many genes and gene products. Microarray technology allows monitoring of gene expression levels for thousands of genes in a single experiment. This is achieved by simultaneously hybridizing two differentially labeled fluorescent cDNA pools to glass slides that contain spots of DNA (Schena et al. (1995) Science 270: 467-70). The US Arabidopsis Functional Genomics Consortium (AFGC) has recently made public the results from such microarray experiments conducted with AFGC chips containing some 10,000 non-redundant ESTs, selected from about 37,000 randomly sequenced ESTs generated from mRNA of different tissues and developmental stages.

The sequences of the ESTs showing at least two-fold increases or decreases in plants treated with 10 μM cadmium compared with untreated plants were identified, compared to the Ceres full length cDNA and genomic sequence databanks, and the equivalent Ceres clones identified. The MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table(s) report(s) the results of this analysis, indicating those Ceres clones which are up or down regulated over controls, thereby indicating the Ceres clones which represent cadmium responsive genes.

Examples of such cadmium responsive genes and gene products are shown in the Reference and Sequence Tables. These genes and/or products are responsible for effects on traits such as plant vigor and seed yield.

While cadmium responsive polynucleotides and gene products can act alone, combinations of these polynucleotides also affect growth and development. Useful combinations

include different cadmium responsive polynucleotides and/or gene products that have similar transcription profiles or similar biological activities, and members of the same or similar biochemical pathways. Whole pathways or segments of pathways are controlled by transcription factor proteins and proteins controlling the activity of signal transduction pathways. Therefore, manipulation of such protein levels is especially useful for altering phenotypes and biochemical activities of plants. In addition, the combination of a cadmium responsive polynucleotide and/or gene product with other environmentally responsive polynucleotides is also useful because of the interactions that exist between, for example, stress and pathogen induced pathways, nutritional pathways and development. Here, in addition to polynucleotides having similar transcription profiles and/or biological activities, useful combinations include polynucleotides that may have different transcription profiles but which participate in common or overlapping pathways.

The MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff Table(s) reports the transcript levels of the experiment (see EXPT ID: Cadium (relating to SMD 7427, SMD 7428)). For transcripts that had higher levels in the samples than the control, a "+" is shown. A "-" is shown for when transcript levels were reduced in root tips as compared to the control. For more experimental detail see the Example section below.

Cadium genes are those sequences that showed differential expression as compared to controls, namely those sequences identified in the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff tables with a "+" or "-" indication

Cadium Genes Identified By Cluster Analyses Of Differential Expression

Cadium Genes Identified By Correlation To Genes That Are Differentially Expressed

As described above, the transcription profiles of genes that act together are well correlated. Applicants not only have identified the genes that are differentially expressed in the microarray experiments, but also have identified the genes that act in concert with them. The MA_clust table indicates groups of genes that have well correlated transcription profiles and therefore participate in the same pathway or network.

A pathway or network of Cadium genes is any group in the MA_clust that comprises a cDNA ID that also appears in Expt ID Cadium (relating to SMD 7427, SMD 7428) of the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table(s).

Cadium Genes Identified By Correlation To Genes That Cause Physiological Consequences

Additionally, the differential expression data and the phenotypic observations can be merged to identify pathways or networks of Cadium genes. A group in the MA_clust is considered a Cadium pathway or network if the group comprises a cDNA ID that also appears in Knock-in or Knock-out tables that causes one or more of the phenotypes described in section above.

Cadium Genes Identified By Amino Acid Sequence Similarity

Cadium genes from other plant species typically encode polypeptides that share amino acid similarity to the sequences encoded by corn and Arabidopsis Cadium genes. Groups of Cadium genes are identified in the Protein Group table. In this table, any protein group that comprises a peptide ID that corresponds to a cDNA ID member of a Cadium pathway or network is a group of proteins that also exhibits Cadium functions/utilities.

Such cadmium responsive genes and gene products can function to either increase or dampen phenotypes or activities either in response to changes in cadmium concentration or in the absence of cadmium fluctuations. Further, promoters of cadmium responsive genes, as described in the Reference tables, for example, are useful to modulate transcription that is induced by cadmium or any of the following phenotypes or biological activities below.

V.K.1. USE OF CADMIUM RESPONSIVE GENES, GENE COMPONENTS AND PRODUCTS TO MODULATE PHENOTYPES

Cadmium responsive genes and gene products are useful to or modulate one or more of the following phenotypes:

- Growth
- Roots

- Initiation and maintenance of cell division
- Stems
- Leaves
- Development
- Mitochondria
- Post-embryonic root meristem development
- Senescence
- Stress Response
- Modulation of Jasmonic Acid and other stress control pathways
- Metabolic detoxification
- Heavy metals
    - Plant and Seed Yield; Fruit Yield

Further, any desired sequence can be transcribed in similar temporal, tissue, or environmentally specific patterns as the cadmium responsive genes when the desired sequence is operably linked to a promoter of a cadmium responsive gene.

To regulate any of the phenotype(s) above, activities of one or more of the cadmium responsive genes or gene products can be modulated and tested by screening for the desired trait. Specifically, the gene, mRNA levels, or protein levels can be altered in a plant utilizing the procedures described herein and the phenotypes can be assayed. As an example, a plant can be transformed according to Bechtold and Pelletier (1998) Methods. Mol. Biol. 82:259-266) and/or screened for variants as in Winkler et al. (1998) Plant Physiol 118: 743-50 and visually inspected for the desired phenotype or metabolically and/or functionally assayed according to Ghoshroy et al. (1998, Plant J 13: 591-602), Citovsky et al. (1998, Plant J 16: 13-20), Clemens et al. (1999, EMBO J 18: 3325-33), Chen et al. (2000, Chemosphere 41: 229-34), Xian and Oliver (1998, Plant Cell 10: 1539-90), Romero-Peurtas et al. (1999, Free Rad Res 31: S25-31), Gaur and Noraho (1995, Biomed Environ Sci 8: 202-10), Thomine et al. (2000, PNAS USA 97: 4991-6), Howden et al. (1995, Plant Physiol 107: 1067-73), Kesseler and Brand (1994, Eur J Biochem 225: 907-22) and Vernoux et al. (2000, Plant Cell 12: 97-110).

V.K.2. USE OF CADMIUM-RESPONSIVE GENES, GENE COMPONENTS AND PRODUCTS TO MODULATE BIOCHEMICAL ACTIVITIES

The activities of one or more of the cadmium responsive genes can be modulated to change biochemical or metabolic activities and/or pathways such as those noted below. Such biological activities can be measured according to the citations included in the Table below:

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
Growth , Differentiation and Development	Root Growth Initiation and maintenance of cell division Resistance to Cadmium-inhibition of root growth	Thomine et al. (2000) PNAS USA 97: 4991-6 Vernoux et al. (2000) Plant Cell 12: 97-110
Metabolism	Cadmium sensing	Howden et al. (1995) Plant Physiol 107: 1067-73
	Cadmium uptake and transport Decreased cadmium transport Phytoremediation	Gaur and Noraho (1995) Biomed Environ Sci 8: 202-10 Thomine et al. (2000) PNAS USA 97: 4991-6
	Inhibition of oxidative phophorylation	Kesseler and Brand (1994) Eur. Biochem 225: 907-22
Plant Defenses	Viral resistance Inhibition of systemic movement of virus Block of viral disease	Ghoshroy et al. (1998) Plant J 13: 591-602
	Detoxification of heavy metals	Clemens et al. (1999) EMBO J 18: 3325-33
	Enhanced stress resistance	Romero-Peurtas et al. (1999) Free Rad Res 31: S25-31
	Cadmium resistance	Xiang and Oliver (1998) Plant

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
	via modulation of jasmonic acid signaling pathway	Cell 10: 1539-90
Signaling	Relief of post-translational gene silencing	Citovsky et al. (1998) Plant J 16: 13-20

Other biological activities that can be modulated by the cadmium responsive genes and gene products are listed in the Reference tables. Assays for detecting such biological activities are described in the Protein Domain table.

Cadmium responsive genes are characteristically differentially transcribed in response to fluctuating cadmium levels or concentrations, whether internal or external to an organism or cell. The MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table(s) report(s) the changes in transcript levels of various cadmium responsive genes following treatment with 10 μM cadmium, relative to untreated plants.

Profiles of some cadmium responsive genes are shown in the Table below together with examples of the kinds of associated biological activities.

TRANSCRIPT LEVELS	TYPE OF GENES	PHYSIOLOGICAL CONSEQUENCES	EXAMPLES OF BIOCHEMICAL ACTIVITY
Up regulated transcripts	Responders to cadmium Application Genes induced by cadmium	Cadmium perception Cadmium uptake and transport Cadmium metabolism Synthesis of secondary metabolites and/or proteins Modulation of cadmium response	Transporters Metabolic enzymes Change in cell membrane structure and potential Kinases and Phosphatases Transcription activators Change in chromatin structure and/or localized

TRANSCRIPT LEVELS	TYPE OF GENES	PHYSIOLOGICAL CONSEQUENCES	EXAMPLES OF BIOCHEMICAL ACTIVITY
		transduction pathways Specific gene transcription initiation Genes involved in inhibiting systemic movement of plant viral RNA Genes involved in post translational gene silencing	DNA topology RNA binding proteins
Down-regulated transcripts	Responders to cadmium Genes repressed by cadmium Genes with discontinued expression or unsTable mRNA in presence of cadmium	Negative regulation of cadmium pathways released Changes in pathways and processes operating in cells Changes in metabolism other than cadmium pathways Genes involved in facilitating systemic movement of plant viral RNA Genes involved in promoting post translational gene silencing	Transcription factors Change in protein structure by phosphorylation (kinases) or Dephosphoryaltion (phosphatases) Change in chromatin structure and/or DNA topology Factors for protein synthesis and degradation Metabolic enzymes RNA binding proteins

Use of Promoters of Cadmium Responsive Genes

Promoters of Cadmium responsive genes are useful for transcription of any desired polynucleotide or plant or non-plant origin. Further, any desired sequence can be transcribed in a similar temporal, tissue, or environmentally specific patterns as the Cadmium responsive genes where the desired sequence is operably linked to a promoter of a Cadmium responsive gene. The protein product of such a polynucleotide is usually synthesized in the same cells, in response to the same stimuli as the protein product of the gene from which the promoter was derived. Such promoter are also useful to produce antisense mRNAs to down-regulate the product of proteins, or to produce sense mRNAs to down-regulate mRNAs via sense suppression.

V.L. DISEASE RESPONSIVE GENES, GENE COMPONENTS AND PRODUCTS

Often growth and yield are limited by the ability of a plant to tolerate stress conditions, including pathogen attack. To combat such conditions, plant cells deploy a battery of inducible defense responses, including the triggering of an oxidative burst and the transcription of pathogenesis-related protein (PR protein) genes. These responses depend on the recognition of a microbial avirulence gene product (avr) by a plant resistance gene product (R), and a series of downstream signaling events leading to transcription- independent and transcription-dependent disease resistance responses. Reactive oxygen species (ROS) such as H_2O_2 and NO from the oxidative burst plays a signaling role, including initiation of the hypersensitive response (HR) and induction of systemic acquired resistance (SAR) to secondary infection by unrelated pathogens. PR proteins are able to degrade the cell walls of invading microorganisms, and phytoalexins are directly microbicidal.

The presence of an avirulent pathogen and/or changes in the concentrations of O_2^-, H_2O_2 and NO in the environment surrounding a plant cell modulate the activities of many genes and, therefore, the levels of many gene products. Examples of tobacco mosaic virus (TMV) responsive genes and gene products, many of them operating through an ROS signaling system, are shown in The Reference and Sequence Tables. These genes and/or products are responsible for effects on traits such as plant vigor and seed yield. The genes were discovered and characterized from a much larger set by experiments designed to find genes whose mRNA products changed in response to application of TMV to plants.

Microarray technology allows monitoring of gene expression levels for thousands of genes in a single experiment. This is achieved by hybridizing labeled fluorescent cDNA pools to glass slides that contain spots of DNA (Schena et al. (1995) Science 270: 467-70). The US Arabidopsis Functional Genomics Consortium (AFGC) has recently made public the results from such microarray experiments conducted with AFGC chips containing some 10,000 non-redundant ESTs, selected from about 37,000 randomly sequenced ESTs generated from mRNA of different tissues and developmental stages.

The sequences of the ESTs showing at least two-fold increases or decreases in response to TMV infection over the non infected controls were identified, compared to the Ceres full length cDNA and genomic sequence databanks, and equivalent Ceres clones identified. The MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table(s) report(s) the

results of this analysis, indicating those Ceres clones which are up or down regulated over controls, thereby indicating the Ceres clones which represent disease responsive genes.

Manipulation of one or more disease responsive gene activities are useful to modulate the biological processes and/or phenotypes listed below. Disease responsive genes and gene products can act alone or in combination. Useful combinations include disease responsive genes and/or gene products with similar transcription profiles, similar biological activities, or members of the same or functionally related biochemical pathways. Whole pathways or segments of pathways are controlled by transcription factor proteins and proteins controlling the activity of signal transduction pathways. Therefore, manipulation of such protein levels is especially useful for altering phenotypes and biochemical activities of plants.

Such disease responsive genes and gene products can function to either increase or dampen the above phenotypes or activities either in response to changes in active oxygen concentration or in the absence of active oxygen fluctuations. The MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff Table(s) reports the transcript levels of the experiment (see EXPT ID: Disease (relating to SMD 7342, SMD 7343)). For transcripts that had higher levels in the samples than the control, a "+" is shown. A "-" is shown for when transcript levels were reduced in root tips as compared to the control. For more experimental detail see the Example section below.

Disease genes are those sequences that showed differential expression as compared to controls, namely those sequences identified in the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff tables with a "+" or "-" indication

Disease Genes Identified By Cluster Analyses Of Differential Expression

Disease Genes Identified By Correlation To Genes That Are Differentially Expressed

As described above, the transcription profiles of genes that act together are well correlated. Applicants not only have identified the genes that are differentially expressed in the microarray experiments, but also have identified the genes that act in concert with them. The MA_clust table indicates groups of genes that have well correlated transcription profiles and therefore participate in the same pathway or network.

A pathway or network of Disease genes is any group in the MA_clust that comprises a cDNA ID that also appears in Expt ID Disease (relating to SMD 7342, SMD 7343) of the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table(s).

Disease Genes Identified By Correlation To Genes That Cause Physiological Consequences

Additionally, the differential expression data and the phenotypic observations can be merged to identify pathways or networks of Disease genes. A group in the MA_clust is considered a Disease pathway or network if the group comprises a cDNA ID that also appears in Knock-in or Knock-out tables that causes one or more of the phenotypes described in section above.

Disease Genes Identified By Amino Acid Sequence Similarity

Disease genes from other plant species typically encode polypeptides that share amino acid similarity to the sequences encoded by corn and Arabidopsis Disease genes. Groups of Disease genes are identified in the Protein Group table. In this table, any protein group that comprises a peptide ID that corresponds to a cDNA ID member of a Disease pathway or network is a group of proteins that also exhibits Disease functions/utilities.

Further, promoters of disease responsive genes, as described in the Reference tables, for example, are useful to modulate transcription that is induced by disease or any of the following phenotypes or biological activities below. Further, any desired sequence can be transcribed in similar temporal, tissue, or enviromentally specific patterns as the disease responsive genes when the desired sequence is operably linked to a promoter of a disease responsive gene.

V.L.1. USE OF DISEASE RESPONSIVE GENES, GENE COMPONENTS AND PRODUCTS TO MODULATE PHENOTYPES

Disease responsive genes and gene products are useful to or modulate one or more of the following phenotypes:

- Pathogen Tolerance and/or Resistance
- Avr/R locus interactions
- Non-Host interactions

- HR
- SAR
- Resistance to bacteria e.g. to Erwinia stewartii, Pseudomonas syringae, Pseudomonas tabaci, Stuart's wilt, etc.
- Resistance to fungi e.g. to downy mildews such as Scleropthora macrospora, Sclerophthora rayissiae, Sclerospora graminicola, Peronosclerospora sorghi, Peronosclerospora philippinensis, Peronosclerospora sacchari, Peronosclerospora maydis; rusts such as Puccinia sorphi, Puccinia polysora, Physopella zeae, etc.; and to other fungal diseases e.g. Cercospora zeae-maydis, Colletotrichum graminicola, Fusarium monoliforme, Exserohilum turcicum, Bipolaris maydis, Phytophthora parasitica, Peronospora tabacina, Septoria, etc.;
- Resistance to viruses or viroids e.g. to tobacco or cucumber mosaic virus, ringspot virus, necrosis virus, pelargonium leaf curl virus, red clover mottle virus, tomato bushy stunt virus, and like viruses;
- Resistance to insecs, such as to aphids e.g. Myzus persicae; to beetles and beetle larvae; to lepidoptera larvae e.g. Heliothus etc.
- Resistance to Nematodes, e.g. Meloidogyne incognita etc
- Local resistance in primary (infected) or secondary (uninfected) leaves
- Stress Tolerance
- Winter Survival
- Cold Tolerance
- Salt tolerance
- Heavy Metal Tolerance, such as Cadmium
- Tolerance to Physical Wounding;
- Increased Organelle Tolerance to Redox Stress, such as in Mitochondria, and chloroplasts
- Cell Death
- Programmed cell death, including death of diseased tissue and during senescence
- Fruit Drop
- Biomass
- Fresh and Dry Weight during any time in plant life, such as maturation

- Number of Flowers, Seeds, Branches, and/or Leaves
- Seed Yield, including Number, Size, Weight, and/or Harvest Index
- Fruit Yield, including Number, Size, Weight, and/or Harvest Index
- Plant Development
- Time to Fruit Maturity
- Cell Wall Strengthening and Reinforcement
- Plant Product Quality
- Paper making quality
- Food additives
- Treatment of Indications modulated by Free Radicals
- Cancer
- Kinds of low molecular weight compounds such as phytoalexins
- Abundance of low molecular weight compounds such as phytoalexins
- Other phenotypes based on gene silencing

To regulate any of the phenotype(s) above, activities of one or more of the disease responsive genes or gene products can be modulated and the plants can be tested by screening for the desired trait. Specifically, the gene, mRNA levels, or protein levels can be altered in a plant utilizing the procedures described herein and the phenotypes can be screened for variants as in Winkler et al. (1998) Plant Physiol 118: 743-50 and assayed , for example, in accordance to Alvarez et al., (1998) Cell 92: 773-784; Halhbrock and Scheel, (1989) Ann. Rev. Plant Physiol. Plant Mol. Biol. 40: 347-369; Lamb et al., (1997) Ann. Rev. Plant Mol. Biol. Plant Physiol. 48: 251-275; Lapwood et al. (1984) Plant Pathol. 33: 13-20; Levine et al. (1996) Curr. Biol. 6: 427-437; McKersie et al., (2000) Plant Physiol. 122: 1427-1437; Olson and Varner (1993) Plant J. 4: 887-892; Pastore et al., (2000), FEBS Lett 470: 88-92; Pastori et al., (1997) Plant Physiol. 113: 411-418; Romero-Puertas et al., (1999) Free Radic. Res. 1999 31 Suppl: S25-31; Shirataki et al., Anticancer Res 20: 423-426 (2000); Wu et al., (1995) Plant Cell 7: 1357-1368.

V.L.2. USE OF DISEASE RESPONSIVE GENES, GENE COMPONENTS AND PRODUCTS TO MODULATE BIOCHEMICAL ACTIVITIES

The activities of one or more of the disease responsive genes can be modulated to change biochemical or metabolic activities and/or pathways such as those noted below. Such biological activities are documented and can be measured according to the citations above and included in the Table below:

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
Resistance to Pathogens	Induction of ROS signaling pathways	Wu et.al.(1995) Plant Cell 7: 1357-68
	Modulation of nitric oxide signaling	Delledonne et al. (1998) Nature 394: 585-588
	Induction of PR proteins, phytoalexins, and defense pathways	Chamnongpol et.al.(1998) Proc. Nat.Acad Sci USA 12;95:5818-23. Davis et al. (1993) Phytochemistry 32: 607-611
	Induction of cellular protectant genes such as glutathione S-transferase (GST) and ascorbate peroxidase	Chen et.al. Plant J. (1996) 10:955-966 Gadea et.al.(1999) Mol Gen Genet 262:212-219 Wu et.al.(1995) Plant Cell 7: 1357-68
	ROS levels following wounding and changes in physical pressure	Orozco-Cardenas and Ryan (1999) Proc.Nat. Acad. Sci. USA 25;96:6553-7. Yahraus et al. (1995) Plant Physiol. 109: 1259-1266

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
	Salicyclic acid levels and signaling	Durner and Klessig (1996) J.Biol.Chem. 271:28492-501
Responses to Wounding	Expression of genes Involved in wound repair and cell division	Legendre et al. (1993) Plant Physiol. 102: 233-240
Responses to Environmental Stress	Expression of genes involved in responses to drought, cold, salt, heavy metals	Shi et al. (2000) Proc. Natl. Acad. Sci. USA 97:6896-6901
Reinforcement of Cell Walls	Modulation of the Production of ExtracTable Proline-Rich Protein	Bradley et al. (1992) Cell 70, 21-30
	Modulation of Lignification	Mansouri et al. (1999) Physiol. Plant 106: 355-362
Programmed Cell Death	Induction of PCD activating genes	Levine et al. (1996) Curr. Biol. 6: 427-437. Reynolds et.al. (1998) Biochem.J. 330:115-20
	Suppression of PCD suppressing genes	Pennell and Lamb (1997) Plant Cell 9, 1157-1168

Other biological activities that can be modulated by the disease responsive genes and their products are listed in the Reference Table. Assays for detecting such biological activities are described in the Protein Domain table.

Disease responsive genes are characteristically differentially transcribed in response to fluctuating levels of disease. The MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table(s)report(s) the changes in transcript levels of various disease responsive genes in the aerial parts of a plant 3 days after the plant was sprayed with a suspension of TMV relative to control plants sprayed with water.

The data from this experiment reveal a number of types of disease responsive genes and gene products, including "early responders," and "delayed responders". Profiles of individual disease responsive genes are shown in the Table below with examples of which associated biological activities are modulated when the activities of one or more such genes vary in plants.

GENE EXPRESSION LEVELS	FUNCTIONAL CATEGORY OF GENE	PHYSIOLOGICAL CONSequence	EXAMPLES OF BIOCHEMICAL ACTIVITY OF GENE PRODUCTS
Upregulated transcripts	Early Responders to Pathogens	ROS Perception and Response	Transcription factors, kinases, phosphatases, GTP-binding proteins (G-proteins), leucine rich repeat proteins (LRRs), transporters, calcium binding proteins, chromatin remodeling proteins
		Initiation of Gene Transcription	Glutathione S-transferase (GST), heat shock proteins, salicylic acid (SA) response pathway proteins, jasmonate response pathway proteins, dehydrins, peroxidases, catalases
	Delayed Responders to Pathogens	Initiation of Defence Gene Transcription	Proteases, pathogen response (PR) proteins, cellulases, chitinases, cutinases, glucanases, other degrading enzymes, calcium channel blockers, phenylalanine ammonia lyase, proteins of defense pathways, cell wall proteins incuding proline rich

			proteins and glycine rich proteins, epoxide hydrolase, methyl transferases
		Activation of cell death pathways	Transcription factors kinases, phosphatases, DNA surveillance proteins, p53, proteases, endonucleases, GTP-binding proteins (G-proteins), leucine rich repeat proteins (LRRs), transporters, calcium binding proteins, mitochondrial and chloroplast energy related proteins, ribosome inactivating proteins
		Initiation of Cellular Protectant Gene Transcription	Reactive oxygen scavenging enzymes, GST, catalase, peroxidase, ascorbate oxidase
Downregulated transcripts	Early responders to pathogens	Negative regulation of pathogen inducible pathways released	Transcription factors, kinases, phosphatases, GTP-binding proteins (G-proteins), leucine rich repeat proteins (LRRs), transporters, calcium binding proteins, chromatin remodelling proteins
	Genes repressed by TMV	Negative regulation of ROS inducible	Transcription factors, kinases, phosphatases, GTP-

		pathways released	binding proteins (G-proteins), leucine rich repeat proteins (LRRs), transporters, calcium binding proteins, chromatin remodelling proteins
	Delayed Responders to Pathogens	Negative regulation of pathogen inducible pathways released	Transcription factors, kinases, phosphatases, GTP-binding proteins (G-proteins), leucine rich repeat proteins (LRRs), transporters, calcium binding proteins, chromatin remodelling proteins
	Genes repressed by TMV	Negative regulation of genes suppressing programmed cell death released	Transcription factors, kinases, phosphatases, GTP-binding proteins (G-proteins), leucine rich repeat proteins (LRRs), transporters, calcium binding proteins, chromatin remodelling proteins

Use of Promoters of Disease Responsive Genes

Promoters of Disease responsive genes are useful for transcription of any desired polynucleotide or plant or non-plant origin. Further, any desired sequence can be transcribed in a similar temporal, tissue, or environmentally specific patterns as the Disease responsive genes where the desired sequence is operably linked to a promoter of a Disease responsive gene. The protein product of such a polynucleotide is usually synthesized in the same cells, in response to the same stimuli as the protein product of the gene from which the promoter was derived. Such

promoter are also useful to produce antisense mRNAs to down-regulate the product of proteins, or to produce sense mRNAs to down-regulate mRNAs via sense suppression.

V.M. DEFENSE (LOL2) RESPONSIVE GENES, GENE COMPONENTS AND PRODUCTS

Often growth and yield are limited by the ability of a plant to tolerate stress conditions, including pathogen attack. To combat such conditions, plant cells deploy a battery of inducible defense responses, including the triggering of an oxidative burst and the transcription of pathogenesis-related protein (PR protein) genes. Reactive oxygen species (ROS) such as H_2O_2 and NO from the oxidative burst play a signaling role, including initiation of the hypersensitive response (HR) and induction of systemic acquired resistance (SAR) to secondary infection by unrelated pathogens. Some PR proteins are able to degrade the cell walls of invading microorganisms, and phytoalexins are directly microbicidal. Other defense related pathways are regulated by salicylic acid (SA) or methyl jasmonate (MeJ).

These responses depend on the recognition of a microbial avirulence gene product (avr) by a plant resistance gene product (R), and a series of downstream signaling events leading to transcription- independent and transcription-dependent disease resistance responses. Current models suggest that R- gene-encoded receptors specifically interact with pathogen-encoded ligands to trigger a signal transduction cascade. Several components include ndr1 and eds1 loci. NDR1, EDS1, PR1, as well as PDF1.2, a MeJ regulated gene and Nim1, a SA regulated gene, are differentially regulated in plants with mutations in the LOL2 gene.

LOL2 shares a novel zinc finger motif with LSD1, a negative regulator of cell death and defense response. Due to an alternative splice site the LOL2 gene encodes two different proteins, one of which contains an additional, putative DNA binding motif. Northern analysis demonstrated that LOL2 transcripts containing the additional DNA binding motif are predominantly upregulated after treatment with both virulent and avirulent Pseudomonas syringae pv maculicola strains. Modulation in this gene can also confer enhanced resistance to virulent and avirulent Peronospora parasitica isolates

Examples of LOL2 responsive genes and gene products are shown in the Reference, Sequence, Protein Group, Protein Group Matrix, MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff and MA_clust tables. These genes and/or products are responsible for effects on traits such as plant vigor, disease resistance, and seed yield. The genes were discovered and characterized from a much larger set by microarray experiments designed to find genes whose mRNA products changed when the LOL2 gene was mutated in plants.

Microarray technology allows monitoring of gene expression levels for thousands of genes in a single experiment. This is achieved by hybridizing labeled fluorescent cDNA pools to glass slides that contain spots of DNA (Schena et al. (1995) Science 270: 467-70). The US Arabidopsis Functional Genomics Consortium (AFGC) has recently made public the results from such microarray experiments conducted with AFGC chips containing some about 10,000 non-redundant ESTs, selected from about 37,000 randomly sequenced ESTs generated from mRNA of different tissues and developmental stages.

The sequences of the ESTs showing at least two-fold increases or decreases in plants with the LOL2 mutation versus wildtype were obtained. Specifically, the plant line lol-2-2 tested, a loss of function mutation. The ESTs were compared to the Ceres full length cDNA and genomic sequence databanks, and equivalent Ceres clones identified. The MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table reports the results of this analysis, indicating those Ceres clones which are up or down regulated over controls, thereby indicating the Ceres clones which represent LOL2 responsive genes.

Manipulation of one or more LOL2 responsive gene activities are useful to modulate the biological processes and/or phenotypes listed below. LOL2 responsive genes and gene products can act alone or in combination. Useful combinations include LOL2 responsive genes and/or gene products with similar transcription profiles, similar biological activities, or members of the same or functionally related biochemical pathways. Whole pathways or segments of pathways are controlled by transcription factor proteins and proteins controlling the activity of signal transduction pathways. Therefore, manipulation of such protein levels is especially useful for altering phenotypes and biochemical activities of plants.

Such LOL2 responsive genes and gene products can function to either increase or dampen the above phenotypes or activities either in response to changes in active LOL2 gene or in the absence. The MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff Table(s) reports the transcript levels of the experiment (see EXPT ID: lol2 (relating to SMD 8031, SMD 8032)). For transcripts that had higher levels in the samples than the control, a "+" is shown. A "-" is shown for when transcript levels were reduced in root tips as compared to the control. For more experimental detail see the Example section below.

Defense genes are those sequences that showed differential expression as compared to controls, namely those sequences identified in the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff tables with a "+" or "-" indication

Defense Genes Identified By Cluster Analyses Of Differential Expression

Defense Genes Identified By Correlation To Genes That Are Differentially Expressed

As described above, the transcription profiles of genes that act together are well correlated. Applicants not only have identified the genes that are differentially expressed in the microarray experiments, but also have identified the genes that act in concert with them. The MA_clust table indicates groups of genes that have well correlated transcription profiles and therefore participate in the same pathway or network.

A pathway or network of Defense genes is any group in the MA_clust that comprises a cDNA ID that also appears in Expt ID lol2 (relating to SMD 8031, SMD 8032) of the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table(s).

Defense Genes Identified By Correlation To Genes That Cause Physiological Consequences

Additionally, the differential expression data and the phenotypic observations can be merged to identify pathways or networks of Defense genes. A group in the MA_clust is considered a Defense pathway or network if the group comprises a cDNA ID that also appears in Knock-in or Knock-out tables that causes one or more of the phenotypes described in section above.

Defense Genes Identified By Amino Acid Sequence Similarity

Defense genes from other plant species typically encode polypeptides that share amino acid similarity to the sequences encoded by corn and Arabidopsis Defense genes. Groups of Defense genes are identified in the Protein Group table. In this table, any protein group that comprises a peptide ID that corresponds to a cDNA ID member of a Defense pathway or network is a group of proteins that also exhibits Defense functions/utilities.

Further, promoters of LOL2 responsive genes, as described in the Reference tables, for example, are useful to modulate transcription that is induced by LOL2 responsive genes or any of the following phenotypes or biological activities below. Further, any desired sequence can be transcribed in similar temporal, tissue, or enviromentally specific patterns as the LOL2 responsive genes when the desired sequence is operably linked to a promoter of a LOL2 responsive gene.

V.M.1. USE OF LOL2 RESPONSIVE GENES, GENE COMPONENTS AND PRODUCTS TO MODULATE PHENOTYPES

Lol2 Responsive Genes And Gene Products Are Useful To Or Modulate One Or More Of The Following Phenotypes:

- Pathogen Tolerance and/or Resistance
- Avr/R locus interactions
- Non-Host interactions
- HR
- SAR, e.g., disease responsive genes acting in conjunction with infection with any of the organisms listed below
- Resistance to bacteria e.g. to Erwinia stewartii, Pseudomonas syringae, Pseudomonas tabaci, Stuart's wilt, etc.
- Resistance to fungi e.g. to downy mildews such as Scleropthora macrospora, Sclerophthora rayissiae, Sclerospora graminicola, Peronosclerospora sorghi, Peronosclerospora philippinensis, Peronosclerospora sacchari, Peronosclerospora maydis; rusts such as Puccinia sorphi, Puccinia polysora, Physopella zeae, etc.; and to other fungal diseases e.g. Cercospora zeae-maydis, Colletotrichum graminicola, Fusarium monoliforme, Exserohilum turcicum, Bipolaris maydis, Phytophthora parasitica, Peronospora tabacina, Septoria, etc.;
- Resistance to viruses or viroids e.g. to tobacco or cucumber mosaic virus, ringspot virus, necrosis virus, pelargonium leaf curl virus, red clover mottle virus, tomato bushy stunt virus, and like viruses;
- Resistance to insecs, such as to aphids e.g. Myzus persicae; to beetles and beetle larvae; to lepidoptera larvae e.g. Heliothus etc.
- Resistance to Nematodes, e.g. Meloidogyne incognita etc
- Local resistance in primary (infected) or secondary (uninfected) leaves
- Stress Tolerance
- Winter Survival
- Cold Tolerance
- Salt tolerance

- Heavy Metal Tolerance, such as Cadmium
- Tolerance to Physical Wounding;
- Increased Organelle Tolerance to Redox Stress, such as in Mitochondria, and chloroplasts
- Cell Death
- Programmed cell death, including death of diseased tissue and during senescence
- Fruit Drop
- Biomass
- Fresh and Dry Weight during any time in plant life, such as maturation
- Number of Flowers, Seeds, Branches, and/or Leaves
- Seed Yield, including Number, Size, Weight, and/or Harvest Index
- Fruit Yield, including Number, Size, Weight, and/or Harvest Index
- Plant Development
- Time to Fruit Maturity
- Cell Wall Strengthening and Reinforcement
- Plant Product Quality
- Paper making quality
- Food additives
- Treatment of Indications modulated by Free Radicals
- Cancer
- Kinds of low molecular weight compounds such as phytoalexins
- Abundance of low molecular weight compounds such as phytoalexins
- Other phenotypes based on gene silencing

To regulate any of the phenotype(s) above, activities of one or more of the LOL2 responsive genes or gene products can be modulated and the plants can be tested by screening for the desired trait. Specifically, the gene, mRNA levels, or protein levels can be altered in a plant utilizing the procedures described herein and the phenotypes can be screened for variants as in Winkler et al. (1998) Plant Physiol 118: 743-50 and assayed , for example, in accordance to Alvarez et al., (1998) Cell 92: 773-784; Halhbrock and Scheel, (1989) Ann. Rev. Plant Physiol. Plant Mol. Biol. 40: 347-369; Lamb et al., (1997) Ann. Rev. Plant Mol. Biol. Plant Physiol. 48:

251-275; Lapwood et al. (1984) Plant Pathol. 33: 13-20; Levine et al. (1996) Curr. Biol. 6: 427-437; McKersie et al., (2000) Plant Physiol. 122: 1427-1437; Olson and Varner (1993) Plant J. 4: 887-892; Pastore et al., (2000), FEBS Lett 470: 88-92; Pastori et al., (1997) Plant Physiol. 113: 411-418; Romero-Puertas et al., (1999) Free Radic. Res. 1999 31 Suppl: S25-31; Shirataki et al., Anticancer Res 20: 423-426 (2000); Wu et al., (1995) Plant Cell 7: 1357-1368.

V.M.2. USE OF DEFENSE RESPONSIVE GENES TO MODULATE BIOCHEMICAL ACTIVITIES

The activities of one or more of the defense (LOL2) responsive genes can be modulated to change biochemical or metabolic activities and/or pathways such as those noted below. Such biological activities are documented and can be measured according to the citations above and included in the Table below:

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
Resistance To Pathogens	Induction Of ROS Signaling Pathways	Wu et.al.(1995) Plant Cell 7: 1357-68
	Modulation Of Nitric Oxide Signaling	Delledonne et al. (1998) Nature 394: 585-588
	Induction Of PR Proteins, Phytoalexins, And Defense Pathways	Chamnongpol et.al.(1998) Proc. Nat.Acad Sci USA 12;95:5818-23. Davis et al. (1993) Phytochemistry 32: 607-611
	Induction Of Cellular Protectant Genes Such As Glutathione S-Transferase (GST) And Ascorbate Peroxidase	Chen et.al. Plant J. (1996) 10:955-966 Gadea et.al.(1999) Mol Gen Genet 262:212-219 Wu et.al.(1995) Plant Cell 7: 1357-68

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
	ROS Levels Following Wounding And Changes In Physical Pressure	Orozco-Cardenas and Ryan (1999) Proc.Nat. Acad. Sci. USA 25;96:6553-7. Yahraus et al. (1995) Plant Physiol. 109: 1259-1266
	Salicyclic Acid Levels And Signaling	Durner and Klessig (1996) J.Biol.Chem. 271:28492-501
Responses To Wounding	Expression Of Genes Involved In Wound Repair And Cell Division	Legendre et al. (1993) Plant Physiol. 102: 233-240
Responses To Environmental Stress	Expression Of Genes Involved In Responses To Drought, Cold, Salt, Heavy Metals	Shi et al. (2000) Proc. Natl. Acad. Sci. USA 97:6896-6901
Reinforcement Of Cell Walls	Modulation Of The Production Of ExtracTable Proline-Rich Protein	Bradley et al. (1992) Cell 70, 21-30
	Modulation Of Lignification	Mansouri et al. (1999) Physiol. Plant 106: 355-362
Programmed Cell Death	Induction Of Pcd Activating Genes	Levine et al. (1996) Curr. Biol. 6: 427-437. Reynolds et.al. (1998) Biochem.J. 330:115-20
	Suppression Of PCD Suppressing Genes	Pennell and Lamb (1997) Plant Cell 9, 1157-1168

Other biological activities that can be modulated by the LOL2 responsive genes and their products are listed in the Reference tables. Assays for detecting such biological activities are described in the Protein Domain table.

LOL2 responsive genes are characteristically differentially transcribed in response to fluctuating levels of disease. MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table reports the changes in transcript levels of various LOL2 responsive genes in the lol-2 line versus control plants.

The data from this experiment reveal a number of types of LOL2 responsive genes and gene products. Profiles of individual LOL2 responsive genes are shown in the Table below with examples of which associated biological activities are modulated when the activities of one or more such genes vary in plants.

GENE EXPRESSION LEVELS	FUNCTIONAL CATEGORY OF GENE	PHYSIOLOGICAL CONSequence	EXAMPLES OF BIOCHEMICAL ACTIVITY OF GENE PRODUCTS
Upregulated transcripts	Early Responders to the LOL2 Mutation	ROS Perception and Response	Transcription factors, kinases, phosphatases, GTP-binding proteins (G-proteins), leucine rich repeat proteins (LRRs), transporters, calcium binding proteins, chromatin remodeling proteins
		Initiation of Gene Transcription	Glutathione S-transferase (GST), heat shock proteins, salicylic acid (SA) response pathway proteins, jasmonate response pathway proteins, dehydrins, peroxidases, catalases
	Delayed Responders to the LOL2	Initiation of Defence Gene Transcription	Proteases, pathogen response (PR) proteins, cellulases, chitinases,

GENE EXPRESSION LEVELS	FUNCTIONAL CATEGORY OF GENE	PHYSIOLOGICAL CONSequence	EXAMPLES OF BIOCHEMICAL ACTIVITY OF GENE PRODUCTS
	Mutation		cutinases, glucanases, other degrading enzymes, calcium channel blockers, phenylalanine ammonia lyase, proteins of defense pathways, cell wall proteins incuding proline rich proteins and glycine rich proteins, epoxide hydrolase, methyl transferases
		Activation of cell death pathways	Transcription factors kinases, phosphatases, DNA surveillance proteins, p53, proteases, endonucleases, GTP-binding proteins (G-proteins), leucine rich repeat proteins (LRRs), transporters, calcium binding proteins, mitochondrial and chloroplast energy related proteins, ribosome inactivating proteins
		Initiation of Cellular Protectant Gene Transcription	Reactive oxygen scavenging enzymes, GST, catalase, peroxidase, ascorbate

GENE EXPRESSION LEVELS	FUNCTIONAL CATEGORY OF GENE	PHYSIOLOGICAL CONSequence	EXAMPLES OF BIOCHEMICAL ACTIVITY OF GENE PRODUCTS
			oxidase
Downregulated transcripts	Early Responders to the LOL2 Mutation	Negative regulation of LOL2 Mutation inducible pathways released	Transcription factors, kinases, phosphatases, GTP-binding proteins (G-proteins), leucine rich repeat proteins (LRRs), transporters, calcium binding proteins, chromatin remodelling proteins
	Genes Repressed by the LOL2 Mutation	Negative regulation of ROS inducible pathways released	Transcription factors, kinases, phosphatases, GTP-binding proteins (G-proteins), leucine rich repeat proteins (LRRs), transporters, calcium binding proteins, chromatin remodelling proteins
	Delayed Responders to the LOL2 Mutation	Negative regulation of LOL2 Mutation inducible pathways released	Transcription factors, kinases, phosphatases, GTP-binding proteins (G-proteins), leucine rich repeat proteins (LRRs), transporters, calcium binding proteins, chromatin remodelling proteins
	Genes	Negative Regulation Of	Transcription Factors,

GENE EXPRESSION LEVELS	FUNCTIONAL CATEGORY OF GENE	PHYSIOLOGICAL CONSequence	EXAMPLES OF BIOCHEMICAL ACTIVITY OF GENE PRODUCTS
	Repressed By The LOL2 Mutation	Genes Suppressing Programmed Cell Death Released	Kinases, Phosphatases, GTP-Binding Proteins (G-Proteins), Leucine Rich Repeat Proteins (Lrrs), Transporters, Calcium Binding Proteins, Chromatin Remodelling Proteins

Use of Promoters of Defense Responsive Genes

Promoters of Defense responsive genes are useful for transcription of any desired polynucleotide or plant or non-plant origin. Further, any desired sequence can be transcribed in a similar temporal, tissue, or environmentally specific patterns as the Defense responsive genes where the desired sequence is operably linked to a promoter of a Defense responsive gene. The protein product of such a polynucleotide is usually synthesized in the same cells, in response to the same stimuli as the protein product of the gene from which the promoter was derived. Such promoter are also useful to produce antisense mRNAs to down-regulate the product of proteins, or to produce sense mRNAs to down-regulate mRNAs via sense suppression.

V.N. IRON RESPONSIVE GENES, GENE COMPONENTS AND PRODUCTS

Iron (Fe) deficiency in humans is the most prevalent nutritional problem worldwide today. Increasing iron availability via diet is a sustainable malnutrition solution for many of the world's nations. One-third of the world's soils, however, are iron deficient. Consequently, to form a food-based solution to iron malnutrition, we need a better understanding of iron uptake, storage and utilization by plants. Furthermore, exposure to non-toxic Fe levels appears to affect inherent plant defense mechanisms. Consequently, exploring the effects of Fe exposure has potential for advances in plant disease resistance in addition to human nutrition.

Microarray technology allows monitoring of gene expression levels for thousands of genes in a single experiment. This is achieved by simultaneously hybridizing two differentially labeled fluorescent FeNA pools to glass slides that contain spots of DNA (Schena et al. (1995) Science 270: 467-70). The Arabidopsis Functional Genomics Consortium (AFGC) has recently made public the results from such microarray experiments conducted with AFGC chips containing 10,000 non-redundant ESTs, selected from 37,000 randomly sequenced ESTs generated from mRNA of different tissues and developmental stages.

The sequences of the ESTs showing at least two-fold increases or decreases over the controls were identified, compared to the Ceres full length FeNA and genomic sequence databanks, and identical Ceres clones identified. MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table reports the results of this analysis, indicating those Ceres clones that are up or down regulated over controls, thereby indicating the Ceres clones which are iron responsive genes.

The MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff Table(s) reports the transcript levels of the experiment (see EXPT ID: Iron (relating to SMD 7114, SMD 7115, SMD 7125)). For transcripts that had higher levels in the samples than the control, a "+" is shown. A "-" is shown for when transcript levels were reduced in root tips as compared to the control. For more experimental detail see the Example section below.

Iron genes are those sequences that showed differential expression as compared to controls, namely those sequences identified in the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff tables with a "+" or "-" indication

Iron Genes Identified By Cluster Analyses Of Differential Expression

Iron Genes Identified By Correlation To Genes That Are Differentially Expressed

As described above, the transcription profiles of genes that act together are well correlated. Applicants not only have identified the genes that are differentially expressed in the microarray experiments, but also have identified the genes that act in concert with them. The MA_clust table indicates groups of genes that have well correlated transcription profiles and therefore participate in the same pathway or network.

A pathway or network of Iron genes is any group in the MA_clust that comprises a cDNA ID that also appears in Expt ID Iron (relating to SMD 7114, SMD 7115, SMD 7125) of the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table(s).

Iron Genes Identified By Correlation To Genes That Cause Physiological Consequences

Additionally, the differential expression data and the phenotypic observations can be merged to identify pathways or networks of Iron genes. A group in the MA_clust is considered a Iron pathway or network if the group comprises a cDNA ID that also appears in Knock-in or Knock-out tables that causes one or more of the phenotypes described in section above.

Iron Genes Identified By Amino Acid Sequence Similarity

Iron genes from other plant species typically encode polypeptides that share amino acid similarity to the sequences encoded by corn and Arabidopsis Iron genes. Groups of Iron genes are identified in the Protein Group table. In this table, any protein group that comprises a peptide ID that corresponds to a cDNA ID member of a Iron pathway or network is a group of proteins that also exhibits Iron functions/utilities.

V.N.1. USE OF IRON RESPONSIVE GENES TO MODULATE PHENOTYPES

Iron responsive genes and gene products are useful to or modulate one or more of the following phenotypes:

- Growth

- Roots
- Root hair formation
- Stems
- Leaves
- Development
- Senescence
- Plant nutrition
- Uptake and assimilation of organic compounds
- Uptake and assimilation of inorganic compounds
- Animal (including human) nutrition
- Improved dietary mineral nutrition
- Stress Response
- Metabolic detoxification
- Heavy metals

To improve any of the phenotype(s) above, activities of one or more of the iron responsive genes or gene products can be modulated and tested by screening for the desired trait. Specifically, the gene, mRNA levels, or protein levels can be altered in a plant utilizing the procedures described herein and the phenotypes can be assayed. As an example, a plant can be transformed according to Bechtold and Pelletier (1998, Methods. Mol. Biol. 82:259-266) and visually inspected for the desired phenotype or metabolically and/or functionally assayed according to Schmidt et al. (2000, Plant Physiol 122:1109-18), Meagher (2000) Current Opinion in Plant Biology 3: 153-62), Deak (1999, Nature Biotechnology (1999, Nature Biotechnology 17: 192-96), Wei and Theil (2000, J. Biol Chem 275: 17488-93) and Vansuyt et al. (1997, FEBS Letters 410: 195-200).

V.N.2. USE OF IRON-RESPONSIVE GENES, GENE COMPONENTS AND PRODUCTS TO MODULATE BIOCHEMICAL ACTIVITIES

The activities of one or more of the iron responsive genes can be modulated to change biochemical or metabolic activities and/or pathways such as those noted below.

Such biological activities can be measured according to the citations included in the Table below:

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
Growth , Differentiation and Development	• Root Growth - Initiation of root hairs	Robinson et al. (1999) Nature 397: 694-97
Metabolisms	• Iron sensing • Iron uptake and transport -decreased iron transport -phytoremediation	Thomine et al. (2000) PNAS USA 97: 4991-6 Thomine et al. (2000) PNAS USA 97: 4991-6 Zhu (1999) Plant Physiol 119: 73-79
Plant Defenses	• Protection from oxidative damage	Deak (1999) Nature Biotechnology 17: 192-6
Signaling	• Specific gene transcription gene silencing	Brand and Perrimon (1993) Development 118: 401-415

Other biological activities that can be modulated by the iron responsive genes and gene products are listed in the REFERENCE Table. Assays for detecting such biological activities are described in the Protein Domain table.

Iron responsive genes are characteristically differentially transcribed in response to fluctuating iron levels or concentrations, whether internal or external to an organism or cell. MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table reports the changes in transcript levels of various iron responsive genes.

The microarray comparison consists of probes prepared from root RNA of A. thaliana (Columbia) seedlings grown under iron-sufficient conditions and seedlings grown under iron-deficient. The data from this experiment reveal a number of types genes and gene products.

Profiles of these different iron responsive genes are shown in the Table below with examples of associated biological activities.

TRANSCRIPT LEVELS	TYPE OF GENES	PHYSIOLOGICAL CONSEQUENCES	EXAMPLES OF BIOCHEMICAL ACTIVITY
Up regulated transcripts	responders to iron application	• Iron perception • Iron uptake and transport • Iron metabolism • Synthesis of secondary metabolites and/or proteins • Modulation of iron response transduction pathways • Specific gene transcription initiation	• Transporters • Metabolic enzymes • Change in cell membrane structure and potential • Kinases and phosphatases • Transcription activators • Change in chromatin structure and/or localized DNA topology
Down-regulated transcripts	responder to iron repressors of iron state of metabolism Genes with discontinued expression or	• Negative regulation of iron pathways • Changes in pathways and processes	• Transcription factors • Change in protein structure by phosphorylation (kinases) or dephosphoryaltion (phosphatases)

TRANSCRIPT LEVELS	TYPE OF GENES	PHYSIOLOGICAL CONSEQUENCES	EXAMPLES OF BIOCHEMICAL ACTIVITY
	unsTable mRNA in presence of iron	operating in cells • Changes in other metabolisms than iron	• Change in chromatin structure and/or DNA topology • Stability of factors for protein synthesis and degradation • Metabolic enzymes

Use of Promoters of Iron Responsive Genes

Promoters of Iron responsive genes are useful for transcription of any desired polynucleotide or plant or non-plant origin. Further, any desired sequence can be transcribed in a similar temporal, tissue, or environmentally specific patterns as the Iron responsive genes where the desired sequence is operably linked to a promoter of a Iron responsive gene. The protein product of such a polynucleotide is usually synthesized in the same cells, in response to the same stimuli as the protein product of the gene from which the promoter was derived. Such promoter are also useful to produce antisense mRNAs to down-regulate the product of proteins, or to produce sense mRNAs to down-regulate mRNAs via sense suppression.

V.O. SHADE RESPONSIVE GENES, GENE COMPONENTS AND PRODUCTS

Plants sense the ratio of Red (R) : Far Red (FR) light in their environment and respond differently to particular ratios. A low R:FR ratio, for example, enhances cell elongation and favors flowering over leaf production. The changes in R:FR ratios mimic and cause the shading response effects in plants. The response of a plant to shade in the canopy structures of agricultural crop fields influences crop yields significantly. Therefore manipulation of genes regulating the shade avoidance responses can improve crop yields. While phytochromes mediate the shade avoidance response, the down-stream factors participating in this pathway are largely unknown. One potential downstream participant, ATHB-2, is a member of the HD-Zip class of transcription factors and shows a strong and rapid response to changes in the R:FR ratio. ATHB-2 overexpressors have a thinner root mass, smaller and fewer leaves and longer hypocotyls and petioles. This elongation arises from longer epidermal and cortical cells, and a decrease in secondary vascular tissues, paralleling the changes observed in wild-type seedlings grown under conditions simulating canopy shade. On the other hand, plants with reduced ATHB-2 expression have a thick root mass and many larger leaves and shorter hypocotyls and petioles. Here, the changes in the hypocotyl result from shorter epidermal and cortical cells and increased proliferation of vascular tissue. Interestingly, application of auxin is able to reverse the root phenotypic consequences of high ATHB-2 levels, restoring the wild-type phenotype. Consequently, given that ATHB-2 is tightly regulated by phytochrome, these data suggest that ATHB-2 may link the auxin and phytochrome pathways in the shade avoidance response pathway.

Changes in R:FR ratios promote changes in gene expression. Microarray technology allows monitoring of gene expression levels for thousands of genes in a single experiment. This is achieved by hybridizing labeled fluorescent cDNA pools to glass slides that contain spots of DNA (Schena et al. (1995) Science 270: 467-70). The US Arabidopsis Functional Genomics Consortium (AFGC) has recently made public the results from such microarray experiments conducted with AFGC chips containing about 10,000 non-redundant ESTs, selected from about37,000 randomly sequenced ESTs generated from mRNA of different tissues and developmental stages.

The sequences of the ESTs showing at least two-fold increases or decreases in plants given 4 hours of FR rich light after growth in high R:FR light compared with the controls of

plants grown in high R:FR light only, were identified, compared to the Ceres full length cDNA and genomic sequence databanks, and equivalent Ceres clones identified. The MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table(s) report(s) the results of this analysis, indicating those Ceres clones which are up or down regulated over controls, thereby indicating the Ceres clones which are shade avoidance responsive genes.

Examples of far red light induced, shade avoidance responsive genes and gene products are shown in the Reference and Sequence Tables. These genes and/or products are responsible for effects on traits such as plant vigor and seed yield.

While far red light, shade avoidance responsive polynucleotides and gene products can act alone, combinations of these polynucleotides also affect growth and development. Useful combinations include different shade avoidance responsive polynucleotides and/or gene products that have similar transcription profiles or similar biological activities, and members of the same or similar biochemical pathways. In addition, the combination of a shade avoidance responsive polynucleotide and/or gene product with another environmentally responsive polynucleotides is also useful because of the interactions that exist between hormone regulated pathways, stress and pathogen induced pathways, nutritional pathways, light induced pathways and development. Here, in addition to polynucleotides having similar transcription profiles and/or biological activities, useful combinations include polynucleotides that may have different transcription profiles but which participate in common or overlapping pathways.

Such far red light induced shade avoidance responsive genes and gene products can function to either increase or dampen the above phenotypes or activities either in response to changes in far red light or in the absence of far red light fluctuations. The MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff Table(s) reports the transcript levels of the experiment (see EXPT ID: Shade (relating to SMD 8130, SMD 7230)). For transcripts that had higher levels in the samples than the control, a "+" is shown. A "-" is shown for when transcript levels were reduced in root tips as compared to the control. For more experimental detail see the Example section below.

Shade genes are those sequences that showed differential expression as compared to controls, namely those sequences identified in the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff tables with a "+" or "-" indication

Shade Genes Identified By Cluster Analyses Of Differential Expression

Shade Genes Identified By Correlation To Genes That Are Differentially Expressed

As described above, the transcription profiles of genes that act together are well correlated. Applicants not only have identified the genes that are differentially expressed in the microarray experiments, but also have identified the genes that act in concert with them. The MA_clust table indicates groups of genes that have well correlated transcription profiles and therefore participate in the same pathway or network.

A pathway or network of Shade genes is any group in the MA_clust that comprises a cDNA ID that also appears in Expt ID Shade (relating to SMD 8130, SMD 7230) of the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table(s).

Shade Genes Identified By Correlation To Genes That Cause Physiological Consequences

Additionally, the differential expression data and the phenotypic observations can be merged to identify pathways or networks of Shade genes. A group in the MA_clust is considered a Shade pathway or network if the group comprises a cDNA ID that also appears in Knock-in or Knock-out tables that causes one or more of the phenotypes described in section above.

Shade Genes Identified By Amino Acid Sequence Similarity

Shade genes from other plant species typically encode polypeptides that share amino acid similarity to the sequences encoded by corn and Arabidopsis Shade genes. Groups of Shade genes are identified in the Protein Group table. In this table, any protein group that comprises a peptide ID that corresponds to a cDNA ID member of a Shade pathway or network is a group of proteins that also exhibits Shade functions/utilities.

Further, promoters of shade avoidance responsive genes, as described in the Reference tables, for example, are useful to modulate transcription that is induced by shade avoidance or any of the following phenotypes or biological activities below. Further, any desired sequence can be transcribed in similar temporal, tissue, or environmentally specific patterns as the shade

avoidance responsive genes when the desired sequence is operably linked to a promoter of a circadian (clock) responsive gene.

V.O.1. USE OF FAR RED RESPONSIVE, SHADE AVOIDANCE RESPONSE GENES TO MODULATE PHENOTYPES

High FR:R, shade avoidance responsive genes and gene products can be used to alter or modulate one or more of the following phenotypes:

- Growth
- Roots
- Elongation
- Lateral root formation
- Stems
- Elongation
- Expansion
- Leaves
- Expansion
- Carotenoid composition
- Development
- Cell
- Growth
- Elongation
- Photosynthetic apparatus
- Efficiency
- Flower
- Flowering time
- Fruit
- Seed
- Dormancy
- Control rate and timing of germination
- Prolongs seed storage and viability
- Inhibition of hydrolytic enzyme synthesis
- Seed and Fruit yield
- Senescence
- Abscission

- Leaf fall
- Flower longevity
- Differentiation
- Vascularization
- Shade (Avoidance) Responses in plant architecture

To regulate any of the phenotype(s) above, activities of one or more of the High FR: R light, shade avoidance responsive genes or gene products can be modulated and the plants tested by screening for the desired trait. Specifically, the gene, mRNA levels, or protein levels can be altered in a plant utilizing the procedures described herein and the phenotypes can be assayed. As an example, a plant can be transformed according to Bechtold and Pelletier (1998, Methods. Mol. Biol. 82:259-266) and/or screened for variants as in Winkler et al. (1998) Plant Physiol 118: 743-50 and visually inspected for the desired phenotype or metabolically and/or functionally assayed according to Carabelli et al. (1996, PNAS USA 93: 3530-3535), Aguirrezabal and Tardieu (1996, J Exp Bot 47: 411-20), Heyer et al. (1995, Plant Physiol 109: 53-61), Garcia-Plazaola et al. (1997, J Exp Bot 48: 1667-74), Schwanz et al. (1996, J Exp Bot 47L 1941-50), Koehne et al. (1999, Biochem Biophys Acta 1412:94-107), Melis (1984, J Cell Biochem 24: 271-85), Steindeler et al. (1999, Development 126: 4235-45), Cruz (1997, J Exp Bot 48: 15-24), Stephanou and Manetas (1997, J Exp Bot 48: 1977-85), Grammatikopoulos et al (1999, J Exp Bot 50:517-21), Krause et al. (1999, Plant Physiol 121: 1349-58), Aukerman et al. (1997, Plant Cell 9: 1317-26), Wagner et al. (1997, Plant Cell 9: 731-43), Weinig (2000) Evolution Int J Org Evolution 54: 124-26), Cocburn et al. (1996, J Exp Bot 47: 647-53), Devlin et al. (1999, Plant Physiol 119: 909-15), Devlin et al. (1998, Plant Cell 10: 1479-87), Finlayson et al. (1998, Plant Physiol 116: 17-25), Morelli and Ruberti (2000, Plant Physiol 122: 621-26), Aphalo et al. (1999, J Exp Bot 50: 1629-34), Sims et al. (1999, J Exp Bot 50: 50: 645-53) and Ballare (1999, Trends Plant Sci 4: 97-102).

V.O.2. USE OF FAR RED LIGHT, SHADE AVOIDANCE RESPONSIVE GENES TO MODULATE BIOCHEMICAL ACTIVITIES

The activities of one or more of the far red light, shade avoidance responsive genes can be modulated to change biochemical or metabolic activities and/or pathways such as those noted below. Such biological activities can be measured according to the citations included in the Table below:

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
Cell Growth and Differentiation	Cell Elongation	Carabelli et al. (1996) PNAS USA 93: 3530-35
	Leaf Expansion	Heyer et al. (1995) Plant Physiol 109: 53-61
Photosynthesis	Development of Photosynthetic Apparatus	Jagtap et al. (1998) J Exp Bot 49: 1715-21 Melis (1984) J Cell Biochem 24: 271-285 McCain (1995) Biophys J 69: 1105-10
	Carotenoid Composition	Garcia-Plazaola et al (1997) J Exp Bot 48: 1667-74
Carbon/Nitrogen Metabolism	Carbon and Nitrogen Assimilation	Cruz (1997) J Exp Bot 48: 15-24
Far red light, shade avoidance response binding by transcription factors		Newton AL, Sharpe BK, Kwan A, Mackay JP, Crossley M. J Biol Chem. 2000May19;275(20):15128-34; Lopez Ribera I, Ruiz-Avila L, Puigdomenech P. Biochem Biophys Res Commun. 1997 Jul 18;236(2):510-6; de Pater S, Greco V, Pham K, Memelink J, Kijne J. Nucleic Acids Res. 1996 Dec 1; 24(23):4624-31.

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
Signaling	UV Light Perception	Stephanou and Manetas (1997) J Exp Bot 48: 1977-85
	Far-red/Red Light Perception	Aukerman et al. (1997) Plant Cell 9: 1317-26 Wagner et al. (1997) Plant Cell 9: 731-43
	Interaction of "Shade Factor" with Ethylene Production/Transduction	Finlayson et al. (1998) Plant Physiol 116: 17-25
	Interaction of "Shade Factor" with Auxin Production/Transduction	Reed et al. (1998) Plant Physiol 118: 1369-78
	Plant to Plant signalling	Sims et al. (1999) J Exp Bot 50: 645-53

Other biological activities that can be modulated by shade avoidance response genes and their products are listed in the REF TABLES. Assays for detecting such biological activities are described in the Protein Domain table.

High FR:R, shade avoidance responsive genes are differentially transcribed in response to high FR:R ratios. The microarray comparison to reveal such genes consisted of probes prepared from RNA isolated from the aerial tissues of A. thaliana (Columbia) two-week old seedlings grown in high R:FR ratios compared to seedlings grown in high R:FR ratios followed by 4 hours of FR-rich light treatment. The data from this experiment reveal a number of types genes and gene products and examples of the classes of genes are given in the Table below.

TRANSCRIPT LEVELS	TYPE OF GENES	PHYSIOLOGICAL CONSEQUENCES	EXAMPLES OF BIOCHEMICAL ACTIVITY
Up regulated transcripts	Responders to high FR:R light ratios Genes induced by high FR:R light ratio	• Far red light perception • Metabolism affected by far red light • Synthesis of secondary metabolites and/or proteins • Modulation of high FR:R light ratio transduction pathways • Specific gene transcription initiation	• Transporters • Metabolic enzymes • Change in cell membrane structure and potential • Kinases and phosphatases • Transcription activators • Change in chromatin structure and/or localized DNA topology • Leaf production factors
Down-regulated transcripts	Responders to high FR:R light ratios Genes repressed by high FR:R light ratio Genes with discontinued expression or	• Changes in pathways and processes operating in cells • Changes in metabolisms other than far red stimulated pathways	• Transcription factors • Change in protein structure by phosphorylation (kinases) or dephosphorylation (phosphatases) • Change in chromatin structure and/or DNA

TRANSCRIPT LEVELS	TYPE OF GENES	PHYSIOLOGICAL CONSEQUENCES	EXAMPLES OF BIOCHEMICAL ACTIVITY
	unsTable mRNA during high FR:R ratio light		topology • Stability of factors for protein synthesis and degradation • Metabolic enzymes • Cell elongation factors • Flowering promotion factors

Use of Promoters of Shade Avoidance Genes

Promoters of Shade Avoidance genes are useful for transcription of any desired polynucleotide or plant or non-plant origin. Further, any desired sequence can be transcribed in a similar temporal, tissue, or environmentally specific patterns as the Shade Avoidance genes where the desired sequence is operably linked to a promoter of a Shade Avoidance gene. The protein product of such a polynucleotide is usually synthesized in the same cells, in response to the same stimuli as the protein product of the gene from which the promoter was derived. Such promoter are also useful to produce antisense mRNAs to down-regulate the product of proteins, or to produce sense mRNAs to down-regulate mRNAs via sense suppression.

V.P. SULFUR RESPONSIVE GENES, GENE COMPONENTS AND PRODUCTS

Sulfur is one of the important macronutrients required by plants. It is taken up from the soil solution by roots as in the form of sulfate anion which higher plants are dependent on to fulfill their nutritional sulfur requirement. After uptake from the soil, sulfate is either accumulated and stored in vacuole or it is assimilated into various organic compounds, e.g. cysteine, glutathione, methionine, etc. Thus, plants also serve as nutritional sulfur sources for animals. Sulfur can be assimilated in one of two ways: it is either incorporated as sulfate in a reaction called sulfation, or it is first reduced to sulfide, the substrate for cysteine synthesis. In plants, majority of sulfur is assimilated in reduced form.

Sulfur comprises a small by vital fraction of the atoms in many protein molecules. As disulfide bridges, the sulfur atoms aid in stabilizing the folded proteins, such cysteine residues. Cys is the first sulfur-containing amino acids, which in proteins form disulfide bonds that may affect the tertiary structures and enzyme activities. This redox balance is mediated by the disulfide/thiol interchange of thioredoxin or glutaredoxin using NADPH as an electron donor. Sulfur can also become sulfhydryl (SH) groups participating in the active sites of some enzymes and some enzymes require the aid of small molecules that contain sulfur. In addition, the machinery of photosynthesis includes some sulfur-containing compounds, such as ferrodoxin.. Thus, sulfate assimilation plays important roles not only in the sulfur nutrition but also in the ubiquitous process that may regulate the biochemical reactions of various metabolic pathways.

Deficiency of sulfur leads to a marked chlorosis in younger leaves, which may become white in color. Other symptoms of sulfur deficiency also include weak stems and reduced growth. Adding sulfur fertilizer to plants can increase root development and a deeper green color of the leaves in sulfur-deficient plants. However, Sulfur is generally sufficient in soils for two reasons: it is a contaminant in potassium and other fertilizers and a product of industrial combustion. Sulfur limitation in plants is thus likely due to the limitation of the uptake and distribution of sulfate in plants. Seven cell type specific sulfate transporter genes have been isolated from Arabidopsis. In sulfate-starved plants, expression of the high-affinity transporter, AtST1-1, is induced in root epidermis and cortex for acquisition of sulfur. The low affinity transporter, AtST2-1 (AST68), accumulates in the root vascular tissue by sulfate starvation for root-to-shoot transport of sulfate. These studies have shown that the whole-plant process of sulfate transport is coordinately regulated by the expression of these 2 sulfate transporter genes

under sulfur limited conditions. Recent studies have proposed that feeding of O-acetylserine, GSH and selenate may regulate the expression of AtST1-1 and AtST2-1 (AST68) in roots either positively or negatively. However, regulatory proteins that may directly control the expression of these genes have not been identified yet.

It has been established that there are regulatory interactions between assimilatory sulfate and nitrate reduction in plants. The two assimilatory pathways are very similar and well coordinated; deficiency for one element was shown to repress the other pathway. The coordination between them should be taken into consideration when one tries to alter one of pathways.

Microarray technology allows monitoring of gene expression levels for thousands of genes in a single experiment. This is achieved by simultaneously hybridizing two differentially labeled fluorescent cDNA pools to glass slides that contain spots of DNA (Schena et al. (1995) Science 270: 467-70). The Arabidopsis Functional Genomics Consortium (AFGC) has recently made public the results from such microarray experiments conducted with AFGC chips containing 10,000 non-redundant ESTs, selected from 37,000 randomly sequenced ESTs generated from mRNA of different tissues and developmental stages.

The sequences of the ESTs showing at least two-fold increases or decreases over the controls were identified, compared to the Ceres full-length cDNA and genomic sequence databanks, and identical Ceres clones identified. MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table reports the results of this analysis, indicating those Ceres clones which are up or down regulated over controls, thereby indicating the Ceres clones which are sulfur response responsive genes.

The MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff Table(s) reports the transcript levels of the experiment (see EXPT ID: Sulfur (relating to SMD 8034, SMD 8035)). For transcripts that had higher levels in the samples than the control, a "+" is shown. A "-" is shown for when transcript levels were reduced in root tips as compared to the control. For more experimental detail see the Example section below.

Sulfur genes are those sequences that showed differential expression as compared to controls, namely those sequences identified in the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff tables with a "+" or "-" indication

Sulfur Genes Identified By Cluster Analyses Of Differential Expression

Sulfur Genes Identified By Correlation To Genes That Are Differentially Expressed

As described above, the transcription profiles of genes that act together are well correlated. Applicants not only have identified the genes that are differentially expressed in the microarray experiments, but also have identified the genes that act in concert with them. The MA_clust table indicates groups of genes that have well correlated transcription profiles and therefore participate in the same pathway or network.

A pathway or network of Sulfur genes is any group in the MA_clust that comprises a cDNA ID that also appears in Expt ID Sulfur (relating to SMD 8034, SMD 8035) of the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table(s).

Sulfur Genes Identified By Correlation To Genes That Cause Physiological Consequences

Additionally, the differential expression data and the phenotypic observations can be merged to identify pathways or networks of Sulfur genes. A group in the MA_clust is considered a Sulfur pathway or network if the group comprises a cDNA ID that also appears in Knock-in or Knock-out tables that causes one or more of the phenotypes described in section above.

Sulfur Genes Identified By Amino Acid Sequence Similarity

Sulfur genes from other plant species typically encode polypeptides that share amino acid similarity to the sequences encoded by corn and Arabidopsis Sulfur genes. Groups of Sulfur genes are identified in the Protein Group table. In this table, any protein group that comprises a peptide ID that corresponds to a cDNA ID member of a Sulfur pathway or network is a group of proteins that also exhibits Sulfur functions/utilities.

V.P.1. USE OF SULFUR RESPONSIVE GENES TO MODULATE PHENOTYPES

Sulfur responsive genes and gene products are useful to or modulate one or more of the following phenotypes:

- Growth
- Roots
- Stems
- Leaves
- Development
- Chloroplasts and Mitochondria
- Fruit Development
- Seed Development
- Seed storage proteins
- Senescence
- Differentiation
- Plastid/chloroplast and mitochondria differentiation
- Protection against oxidative damage
- regulation of enzymes via redox control by thiol groups
- Metabolic detoxification
- Photosynthesis
- Carbon dioxide fixation

To improve any of the phenotype(s) above, activities of one or more of the sulfur responsive genes or gene products can be modulated and tested by screening for the desired trait. Specifically, the gene, mRNA levels, or protein levels can be altered in a plant utilizing the procedures described herein and the phenotypes can be assayed. As an example, a plant can be transformed according to Bechtold and Pelletier (1998, Methods. Mol. Biol. 82:259-266) and visually inspected for the desired phenotype or metabolically and/or functionally assayed according to Saito et al. (1994, Plant Physiol. 106: 887-95), Takahashi et al (1997, Proc. Natl. Acad. Sci. USA 94: 11102-07) and Koprivova et al. (2000, Plant Physiol. 122: 737-46).

V.P.2. USE OF SULFUR-RESPONSIVE GENES, GENE COMPONENTS AND PRODUCTS TO MODULATE BIOCHEMICAL ACTIVITIES

The activities of one or more of the sulfur responsive genes can be modulated to change biochemical or metabolic activities and/or pathways such as those noted below. Such biological activities can be measured according to the citations included in the Table below:

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
Growth, Differentiation and Development	• Root • Leaf • Stem • Chloroplast/Mitochondria development/differentiation	Klein and Klein (1988) Mineral Nutrition, In CM Wilson and J Gregory, eds Fundamentals of Plant Science. Harper and Row Publishers, Inc., NY, p163 Rost et al. (1984) The Absorption and Transport System, In R Bem, ed, Botany-A Brief Introduction to Plant Biology. John Wiley and Sons, NY, p96. Huluigue et al. (2000) Biochem Biophys Res Commun 271: 380-5 Kapazoglou et al. (2000) Eur J Biochem 267: 352-60
	• Seed storage protein synthesis	Kim et al. (1999) 209: 282-9
Metabolisms	• Sulfate uptake and transport	Takahashi et al. (1997) Proc Natl Acad Sci USA 94: 11102-07
	• Cysteine Biosynthesis	Saito et al. (1992) Proc Natl Acad Sci USA 89: 8078-82 Hesse et al. (1999) Amino Acids 16: 113-31
	• Methionine biosynthesis	Bourgis et al. (1999) Plant Cell 11: 1485-98
	• Carbon dioxide fixation in photosynthesis	Buchana (1991) Arch Biochem Biophys 288: 1-9

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
	• Thioredoxin reduction	Leustek and Saito (1999) Plant Phyiol 120: 637-43 Mamedova et al. (1999) FEBS Lett 462: 421-4
	• Nitrogen metabolism	Koprivova et al. (2000) Plant Physiol. 122: 737-46 Yamaguchi et al. (1999) Biosci Biotechnol Biochem 63: 762-6
Plant Defenses	• Reduction of oxidative stress – oxygen metabolism and reactive oxygen species	May et al. (1998) J Expt Bio 49: 649-67 Kreuz et al. (1996) Plant Physiol 111: 349-53
	• Detoxification of toxins, xenobiotics and heavy metals	Zhao et al. (1998) Plant Cell 10: 359-70
	• Defense against pathogens or microbes	Kyung and Fleming (1997) J Food Prot 60: 67-71
	• Disease prevention by secondary sulfur-containing compounds	Fahey et al. (1997) Proc Natl Acad Sci USA 94: 10367-72
	• Activation of kinases and phosphatases	Davis et al. (1999) Plant Cell 11: 1179-90

Other biological activities that can be modulated by the sulfur responsive genes and gene products are listed in the REFERENCE Table. Assays for detecting such biological activities are described in the Protein Domain table.

Sulfur responsive genes are characteristically differentially transcribed in response to fluctuating sulfur levels or concentrations, whether internal or external to an organism or cell.

MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table reports the changes in transcript levels of various sulfur responsive genes.

Profiles of these different sulfur responsive genes are shown in the Table below with examples of associated biological activities.

TRANSCRIPT LEVELS	TYPE OF GENES	PHYSIOLOGICAL CONSEQUENCES	EXAMPLES OF BIOCHEMICAL ACTIVITY
Up regulated transcripts	Responders to sulfur Application	• Sulfur perception • Sulfur uptake and transport • Sulfur metabolism • Synthesis of secondary metabolites and/or proteins • Modulation of sulfur response transduction pathways • Specific gene transcription initiation	• Transporters • Metabolic enzymes • Change in cell membrane structure and potential • Kinases and phosphatases • Transcription activators • Change in chromatin structure and/or localized DNA topology • Redox control
Down-regulated transcripts	responder to sulfur repressors of sulfur state of metabolism Genes with discontinued expression or	• Negative regulation of sulfur pathways • Changes in pathways and processes	• Transcription factors • Change in protein structure by phosphorylation (kinases) or dephosphoryaltion (phosphatases)

TRANSCRIPT LEVELS	TYPE OF GENES	PHYSIOLOGICAL CONSEQUENCES	EXAMPLES OF BIOCHEMICAL ACTIVITY
	unsTable mRNA in presence of sulfur	operating in cells • Changes in other metabolisms than sulfur	• Change in chromatin structure and/or DNA topology • Stability of factors for protein synthesis and degradation • Metabolic enzymes

Use of Promoters of Sulfur Responsive Genes

Promoters of Sulfur responsive genes are useful for transcription of any desired polynucleotide or plant or non-plant origin. Further, any desired sequence can be transcribed in a similar temporal, tissue, or environmentally specific patterns as the Sulfur responsive genes where the desired sequence is operably linked to a promoter of a Sulfur responsive gene. The protein product of such a polynucleotide is usually synthesized in the same cells, in response to the same stimuli as the protein product of the gene from which the promoter was derived. Such promoter are also useful to produce antisense mRNAs to down-regulate the product of proteins, or to produce sense mRNAs to down-regulate mRNAs via sense suppression.

V.Q. ZINC RESPONSIVE GENES, GENE COMPONENTS AND PRODUCTS

Phytoremediation of soils contaminated with toxic levels of heavy metals requires the understanding of plant metal transport and tolerance. The numerous Arabidopsis thaliana studies have given scientists the potential for dissection and elucidation of plant micronutrient/heavy metal uptake and accumulation pathways. It has been shown altered regulation of ZNT1, a Zn/Cd transporter, contributes to high Zn uptake. Isolation and characterization of Zn/Cd hyperaccumulation genes may allow expression in higher biomass plant species for efficient contaminated soil clean up. Identification of additional Zn transport, tolerance and nutrition-related genes involved in heavy metal accumulation will enable manipulation of increased uptake (for phytoremediation) as well as limitation of uptake or leak pathways that contribute to toxicity in crop plants. Additionally, Zn-binding ligands involved in Zn homeostasis or tolerance may be identified, as well as factors affecting the activity or expression of Zn binding transcription factors. Gene products acting in concert to effect Zn uptake, which would not have been identified in complementation experiments, including multimeric transporter proteins, could also be identified.

Microarray technology allows monitoring of gene expression levels for thousands of genes in a single experiment. This is achieved by simultaneously hybridizing two differentially labeled fluorescent cDNA pools to glass slides that contain spots of DNA (Schena et al. (1995) Science 270: 467-70). The Arabidopsis Functional Genomics Consortium (AFGC) has recently made public the results from such microarray experiments conducted with AFGC chips containing 10,000 non-redundant ESTs, selected from 37,000 randomly sequenced ESTs generated from mRNA of different tissues and developmental stages.

The sequences of the ESTs showing at least two-fold increases or decreases over the controls were identified, compared to the Ceres full-length cDNA and genomic sequence databanks, and identical Ceres clones identified. The Zn response information was then used in conjunction with the existing annotation to attribute biological function or utility to the full-length cDNA and corresponding genomic sequence.

The MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff Table(s) reports the transcript levels of the experiment (see EXPT ID: Zinc (relating to SMD 7310, SMD 7311)). For transcripts that had higher levels in the samples than the control, a "+" is shown. A

"-" is shown for when transcript levels were reduced in root tips as compared to the control. For more experimental detail see the Example section below.

Zinc genes are those sequences that showed differential expression as compared to controls, namely those sequences identified in the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff tables with a "+" or "-" indication

Zinc Genes Identified By Cluster Analyses Of Differential Expression

Zinc Genes Identified By Correlation To Genes That Are Differentially Expressed

As described above, the transcription profiles of genes that act together are well correlated. Applicants not only have identified the genes that are differentially expressed in the microarray experiments, but also have identified the genes that act in concert with them. The MA_clust table indicates groups of genes that have well correlated transcription profiles and therefore participate in the same pathway or network.

A pathway or network of Zinc genes is any group in the MA_clust that comprises a cDNA ID that also appears in Expt ID Zinc (relating to SMD 7310, SMD 7311) of the MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff table(s).

Zinc Genes Identified By Correlation To Genes That Cause Physiological Consequences

Additionally, the differential expression data and the phenotypic observations can be merged to identify pathways or networks of Zinc genes. A group in the MA_clust is considered a Zinc pathway or network if the group comprises a cDNA ID that also appears in Knock-in or Knock-out tables that causes one or more of the phenotypes described in section above.

Zinc Genes Identified By Amino Acid Sequence Similarity

Zinc genes from other plant species typically encode polypeptides that share amino acid similarity to the sequences encoded by corn and Arabidopsis Zinc genes. Groups of Zinc genes are identified in the Protein Group table. In this table, any protein group that comprises a peptide ID that corresponds to a cDNA ID member of a Zinc pathway or network is a group of proteins that also exhibits Zinc functions/utilities.

V.Q.1. USE OF ZN TRANSPORT, TOLERANCE AND NUTRITION-RELATED GENES TO MODULATE PHENOTYPES

Changes in zinc concentration in the surrounding environment or in contact with a plant results in modulation of many genes and gene products. Examples of such zinc responsive genes and gene products are shown in the Reference, Sequence tables, Protein Group, Protein Group Matrix, MA_diff and/or AFLP_diff MA_diff and/or AFLP_diff and/or AFLP_diff, and MA_clust tables. These genes and/or products are responsible for effects on traits such as plant vigor and seed yield.

While zinc responsive polynucleotides and gene products can act alone, combinations of these polynucleotides also affect growth and development. Useful combinations include different zinc responsive polynucleotides and/or gene products that have similar transcription profiles or similar biological activities, and members of the same or similar biochemical pathways. In addition, the combination of a zinc responsive polynucleotide and/or gene product with another environmentally responsive polynucleotide is also useful because of the interactions that exist between hormone regulated pathways, stress pathways, nutritional pathways and development. Here, in addition to polynucleotides having similar transcription profiles and/or biological activities, useful combinations include polynucleotides that may have different transcription profiles but which participate in a common pathway.

Such zinc responsive genes and gene products can function to either increase or dampen the above phenotypes or activities either

- in response to changes in zinc concentration or
- in the absence of zinc fluctuations.

Zn transport, tolerance and nutrition-related genes and gene products can be used to alter or modulate one or more of the following phenotypes:

- Zn Uptake
- Transport of Zn or other heavy metals into roots
- Epidermal/cortical uptake
- Xylem loading
- Zn compartmentation
- Xylem unloading

- Phloem loading
- Efflux from cells to apoplast
- Sequestration in vacuoles/subcellular compartments.
- Zn tolerance
- Chelation of Zn
- Transport of Zn
- Metabolic and transcriptional control
- Activity of Zn binding enzymes
- Activity of Zn binding transcription factors

To improve any of the phenotype(s) above, activities of one or more of the Zn transport, tolerance and nutrition-related genes or gene products can be modulated and the plants can be tested by screening for the desired trait. Specifically, the gene, mRNA levels, or protein levels can be altered in a plant utilizing the procedures described herein and the phenotypes can be assayed, for example, in accordance to Lasat MM, Pence NS, Garvin DF, Ebbs SD, Kochian LV. J Exp Bot. 2000 Jan;51(342):71-9; Grotz N, Fox T, Connolly E, Park W, Guerinot ML, Eide D. Proc Natl Acad Sci U S A. 1998 Jun 9;95(12):7220-4; Crowder MW, Maiti MK, Banovic L, Makaroff CA. FEBS Lett. 1997 Dec 1;418(3):351-4; Hart JJ, Norvell WA, Welch RM, Sullivan LA, Kochian LV. Plant Physiol. 1998 Sep;118(1):219-26.

V.Q.2. USE OF ZN TRANSPORT, TOLERANCE AND NUTRITION-RELATED GENES TO MODULATE BIOCHEMICAL ACTIVITIES

Alternatively, the activities of one or more of the zinc responsive genes can be modulated to change biochemical or metabolic activities and/or pathways such as those noted below. Such biological activities can be measured according to the citations included in the Table below:

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
Zn Uptake and Assimilation	Zn Influx	Lasat MM, Pence NS, Garvin DF, Ebbs SD, Kochian LV. J Exp Bot. 2000 Jan;51(342):71-9.
	Zn compartmentation	Hart JJ, Norvell WA, Welch RM, Sullivan LA, Kochian LV. Plant Physiol. 1998 Sep;118(1):219-26.
Zn binding by metabolic enzymes		Crowder MW, Maiti MK, Banovic L, Makaroff CA.FEBS Lett. 1997 Dec 1;418(3):351-4; Kenzior AL, Folk WR. FEBS Lett. 1998 Dec 4;440(3):425-9.
Zn binding by transcription factors		Newton AL, Sharpe BK, Kwan A, Mackay JP, Crossley M. J Biol Chem. 2000May19;275(20):15128-34; Lopez Ribera I, Ruiz-Avila L, Puigdomenech P. Biochem Biophys Res Commun. 1997 Jul 18;236(2):510-6; de Pater S, Greco V, Pham K, Memelink J, Kijne J. Nucleic Acids Res. 1996 Dec 1;24(23):4624-31.

PROCESS	BIOCHEMICAL OR METABOLIC ACTIVITIES AND/OR PATHWAYS	CITATIONS INCLUDING ASSAYS
Synthesis of proteins to chelate Zn and other metals		Schafer HJ, Greiner S, Rausch T, Haag-Kerwer A. FEBS Lett. 1997 Mar 10;404(2-3):216-20. Rauser WE.Cell Biochem Biophys. 1999;31(1):19-48.
Synthesis of metabolites to chelate Zn and other metals		Rauser WE.Cell Biochem Biophys. 1999;31(1):19-48.

Other biological activities that can be modulated by Zn transport, tolerance and nutrition-related genes and their products are listed in the Reference tables. Assays for detecting such biological activities are described in the Protein Domain table.

Zn transport, tolerance and nutrition-related genes are differentially transcribed in response to low Zn concentrations. The microarray comparison consists of probes prepared from root RNA of A. thaliana (Columbia) seedlings hydroponically grown in complete nutrient medium (control) and Zn deficient seedlings grown in -Zn nutrient medium (experimental). The data from this experiment reveal a number of types genes and gene products. MA_diff and/or AFLP_diff table reports the changes in transcript levels of various zinc responsive genes in entire seedlings at 1 and 6 hours after a plant was sprayed with a Hoagland's solution enriched with zinc as compared to seedlings sprayed with Hoagland's solution only.

The data from this time course can be used to identify a number of types of zinc responsive genes and gene products, including "early responding," "high zinc responders," "repressors of zinc deprivation pathways" and "zinc deprivation responders." Profiles of these different zinc responsive genes are shown in the Table below with examples of associated biological activities.

TRANSCRIPT LEVELS	TYPE OF GENE	PHYSIOLOGICAL CONSequence	EXAMPLES OF BIOCHEMICAL ACTIVITY
Upregulated transcripts (level at 1 hour ≅ 6 hours) (level at 1 hour > 6 hours)	Early responders to Zinc Zinc Deprivation Responders	- Zinc Perception -Zinc Uptake - Modulation of Zinc Response Transduction Pathways - Specific Gene Transcription Initiation -Repression of Pathways to Optimize Zinc Response Pathways	-Transcription Factors -Transporters -Inhibit Transport of Zinc -Degradation
Level at 1 hour < 6 hours	Delayed Zinc Responders Repressor of Zinc Deprivation Pathways	Negative Regulation of Zinc Pathways	-Zinc Metabolic Pathways
Down Regulated transcripts (Level at 1 hour ≅ 6 hours) (Level at 6 hours > 1 hour)	Early responder repressors of Zinc utilization Pathways	Negative Regulators of Zinc Utilization Pathways	Suppressing Zinc Requiring processes
Level at 1 hour > 6 hours	Genes with discontinued expression or unsTable mRNA following Zinc uptake	Changes in pathways and processes operating I cells	

Use of Promoters of Zinc Responsive Genes

Promoters of Zinc responsive genes are useful for transcription of any desired polynucleotide or plant or non-plant origin. Further, any desired sequence can be transcribed in a similar temporal, tissue, or environmentally specific patterns as the Zinc responsive genes where the desired sequence is operably linked to a promoter of a Zinc responsive gene. The protein product of such a polynucleotide is usually synthesized in the same cells, in response to the same stimuli as the protein product of the gene from which the promoter was derived. Such promoter are also useful to produce antisense mRNAs to down-regulate the product of proteins, or to produce sense mRNAs to down-regulate mRNAs via sense suppression.

VI. UTILITIES OF PARTICULAR INTEREST

Genes capable of modulating the phenotypes in the following table are useful produce the associated utilities in the table. Such genes can be identified by their cDNA ID number in the Knock-in and Knock-out tables.

Phenotype Modulated by a Gene		**Utilities**
Leaf shape	**Cordate**	decrease wind opacity,
	Cup-shaped	decrease lodging (plant fall over),
	Curled	increase biomass by making larger or different shaped leaves,
	Laceolate	improve the efficiency of mechanical harvesting,
	Lobed	decrease transpiration for better drought tolerance,
	Oval	changing leaf shape to collect and absorb water,
	Ovate	modulation of canopy structure and shading for altered irradiance close to the ground,
	Serrate	enhanced uptake of pesticides (herbicides, fungicides, etc),
	Trident	creation of ornamental leaf shapes,
	Undulate	increase resistance to pathogens by decreasing amount of water that collects on leaves,
	Vertically Oblong	change proporation of cell types in the leaves for enhanced photosynthesis, decreased transpiration, and enhanced
	Other Shapes	accumulation of desirable compounds including secondary metabolites in specialized cells,
		decrease insect feeding,
	Long petioles	decrease wind opacity,
	Short petioles	decrease lodging (plant fall over),
		increase biomass by better positioning of the leaf blade,
		decrease insect feeding,
		decrease transpiration for better drought tolerance,
		position leaves most effectively for

Phenotype Modulated by a Gene		Utilities
		photosynthetic efficiency
	Fused	ornamental applications to make distinctive plants,
Reduced fertility		
	Short siliques	increase or decrease the number of seeds in a fruit, increasing fruit size, modulating fruit shape to better fit harvesting or packaging requirements, useful for controlling dehisence and seed scatter
	Reduced fertility **Sterility**	useful in hybrid breeding programs, increasing fruit size, production of seedless fruit, useful as targets for gametocides, modulating fruit shape to better fit harvesting or packaging requirements, useful for controlling dehisence and seed scatter
	Flower size	useful for edible flowers useful for flower derived products such as fragrances useful for modulating seed size and number in combination with seed-specific genes value in the ornamental industry
Stature	**Large** **Small** **Dwarfs**	increasing or decreasing plant biomass, optimizing plant stature to increase yield under various diverse environmental conditions, e.g., when water or nutrients are limiting, decreasing lodging, increasing fruit number and size, controlling shading and canopy effects
Meristems		Change plant architecture,

Phenotype Modulated by a Gene		Utilities
		increase or decrease number of leaves as well as change the types of leaves to increase biomass, improve photosynthetic efficiency, create new varieties of ornamental plants with enhanced leaf design, preventing flowering to opimize vegetative growth, control of apical dominace, increase or decrease flowering time to fit season, water or fertilizer schedules, change arrangement of leaves on the stem (phyllotaxy) to optimize plant density, decrease insect feeding, or decrease pathogen infection, increase number of trichome/glandular trichome producing leaves targets for herbicides, generate ectopic meristems and ectopic growth of vegetative and floral tissues and seeds and fruits
Stem	**Strong** **Weak**	modify lignin content/composition for creation of harder woods or reduce difficulty/costs in pulping for paper production or increase digestibility of forage crops, decrease lodging, modify cell wall polysaccharides in stems and fruits for improved texture and nutrition. increase biomass
	Late/Early Bolting	Break the need for long vernalization of vernalization-dependent crops, e.g., winter wheat, thereby increasing yield decrease or increase generaton time increase biomass
Lethals	**Embryo-lethal**	produce seedless fruit, use as herbicide targets

Phenotype Modulated by a Gene		**Utilities**
	Embryo-defective	produce seedless fruit, use as herbicide targets
	Seedling	use as herbicide targets, useful for metabolic engineering,
	Pigment-lethals	use as herbicide targets, increase photosynthetic efficiency
Pigment	**Dark Green**	Increase nutritional value, enhanced photosynthesis and carbon dioxide combustion and therefore increase plant vigor and biomass, enhanced photosynthetic efficiency and therefore increase plant vigor and biomass, prolong vegetative development, enhanced protection against pathogens,
	YGV1	Useful as targets for herbicides, increase photosynthetic efficiency and therefore increase plant vigor and biomass,
	YGV2	Useful as targets for herbicides, control of change from embryonic to adult organs, increase metabolic efficiency, increase photosynthetic efficiency and therefore increased plant vigor and biomass,
	YGV3	Useful as targets for herbicides, nitrogen sensing/uptake/usage, increase metabolic efficiency and therefore increased plant vigor and biomass,
	Interveinal chlorosis	to increase photosynthetic efficiency and therefore increase plant vigor and biomass to increase or decrease nitrogen transport and therefore increase plant vigor and

Phenotype Modulated by a Gene		Utilities
		biomass use as herbicide targets increase metabolic efficiency,
Roots	**Short (primary root)**	to access water from rainfall, to access rhizobia spray application, for anaerobic soils, useful to facilitate harvest of root crops,
	Thick (primary root)	useful for increasing biomass of root crops, for preventing plants dislodging during picking and harvesting, as root grafts, for animal feeds
	Branching (primary root)	modulation allows betters access to water, minerals, fertilizers, rhizobia prevent soil erosion,s increasing root biomass decrease root lodging,
	Long (lateral roots)	modulation allows improved access to water, nutrients, fertilizer, rhizobia, prevent soil erosion increase root biomass decrease root lodging modulation allows control on the depth of root growth in soil to access water and nutriennts modulation allows hormonal control of root growth and development (size)
	Agravitropic	modulation allows control on the depth of root growth in soil
	Curling (primary root)	modulation allows hormonal control of root growth and development (size) useful in anaerobic soils in allowing roots to stay close to surface harvesting of root crops
	Poor germination	

Phenotype Modulated by a Gene		Utilities
Trichome	**Reduced Number** **Glabrous**	Genes useful for decreasing transpiration, increased production of glandular trichomes for oil or other secreted chemicals of value,
	Increased Number	use as deterrent for insect herbivory and ovipostion modulation will increase resistance to UV light,
Wax mutants		decrease insect herbivory and oviposition, compostion changes for the cosmetics industry, decrease transpiration, provide pathogen resistance, UV protection, modulation of leaf runoff properties and improved access for herbicides and fertilizers
Cotyledons		modulation of seeds structure in legumes, increase nutritional value, improve seedling competion under field conditions,
Seeds	**Transparent testa**	genes useful for metabolic engineering anthocyanin and flavonoid pathways
	Light **Dark**	improved nutritional content
Flowers	**Other**	decrease petal abscission decrease pod shattering
Hypocotysl	**Long**	to improve germination rates to improve plant survivability
	Short	to improve germination rates to improve plant survivability

VII. ENHANCED FOODS

Animals require external supplies of amino acids that they cannot synthesize themselves. Also, some amino acids are required in larger quantities. The nutritional values of plants for animals and humans can thus be modified by regulating the amounts of the constituent amino acids that occur as free amino acids or in proteins. For instance, higher levels of lysine and/or methionine would enhance the nutritional value of corn seed. Applicants herein provide several methods for modulating the amino acid content:

(1) expressing a naturally occurring protein that has a high percentage of the desired amino acid(s);

(2) expressing a modified or synthetic coding sequence that has an enhanced percentage of the desired amino acids; or

(3) expressing the protein(s) that are capable of synthesizing more of the desired amino acids.

A specific example is expressing proteins with enhanced, for example, methionine content, preferentially in a corn or cereal seed used for animal nutrition or in the parts of plants used for nutritional purposes.

A protein is considered to have a high percentage of an amino acid if the amount of the desired amino acid is at least 1% of the total number of residues in a protein; more preferably 2% or greater. Amino acids of particular interest are tryptophan, lysine, methionine, phenylalanine, threonine leucine, valine, and isoleucine. Examples of naturally occurring proteins with a high percentage of any one of the amino acid of particular interest are listed in the Enhanced Amino Table.

The sequence(s) encoding the selected protein(s) are operably linked to a promoter and other regulatory sequences and transformed into a plant as described below. The promoter is chosen optimally for promoting the desired level of expression of the protein in the selected organ e.g. a promoter highly functional in seeds. Modifications may be made to the sequence encoding the protein to ensure protein transport into, for example, organelles or storage bodies or its accumulation in the organ. Such modifications may include addition of signal sequences at or near the N terminus and amino acid residues to modify protein stability or appropriate

glycosylation. Other modifications may be made to the transcribed nucleic acid sequence to enhance the stability or translatability of the mRNA, in order to ensure accumulation of more of the desired protein. Suitable versions of the gene construct and transgenic plants are selected on the basis of, for example, the improved amino acid content and nutritional value measured by standard biochemical tests and animal feeding trials.

VIII. USE OF NOVEL GENES TO FACILITATE EXPLOITATION OF PLANTS AS FACTORIES FOR THE SYNTHESIS OF VALUBLE MOLECULES

Plants and their constituent cells, tissues, and organs are factories that manufacture small organic molecules such as sugars, amino acids, fatty acids, vitamins, etc., as well as macromolecules such as proteins, nucleic acids, oils/fats and carbohydrates. Plants have long been a source of pharmaceutically beneficial chemicals; particularly, the secondary metabolites and hormone-related molecules synthesized by plants. Plants can also be used as factories to produce carbohydrates or lipids that comprises a carbon backbone useful as precursors of plastics, fiber, fuel, paper, pulp, rubber, solvents, lubricants, construction materials, detergents, and other cleaning materials. Plants can also generate other compounds that are of economic value, such as dyes, flavors, and fragrances. Both the intermediates as well as the end-products of plant bio-synthetic pathways have been found useful.

With the polynucleotides and polypeptides of the instant invention, modification of both in-vitro and in-vivo synthesis of such products is possible. One method of increasing the amount of either the intermediates or the end-products synthesized in a cell is to increase the expression of one or more proteins in the synthesis pathway as discussed below. Another method of increasing production of an intermediate is to inhibit expression of protein(s) that synthesize the end-product from the intermediate. Levels of end-products and intermediates can also be modified by changing the levels of enzymes that specifically change or degrade them. The kinds of molecules made can be also be modified by changing the genes encoding specific enzymes performing reactions at specific steps of the biosynthetic pathway. These genes can be from the same or a different organism. The molecular structures in the biosynthetic pathways can thus be modified or diverted into different branches of a pathway to make novel end-products.

Novel genes comprising selected promoters and sequences encoding enzymes are transformed into the selected plant to modify the levels, composition and/or structure of, without limitation:

- Terpenoids
- Alkaloids
- Hormones, including brassinosteriods
- Flavonoids

- Steroids
- Vitamins such as
  - Retinol
  - Riboflavin
  - Thiamine
- Caffeine
- Morphine and other alkaloids
- Peptides and amino acid synthesis
- Antioxidants
- Starches and lipids
- Fatty acids
- Fructose, mannose and other sugars
- Glycerolipid
- Citric acid
- Lignin
- Flavors
- Fragrances
- Essential oils
- Colors or dyes
- Gum
- Gels
- Waxes

The modifications are made by designing one or more novel genes per application comprising promoters, to ensure production of the enzyme(s) in the relevant cells, in the right amount, and polynucleotides encoding the relevant enzyme. The promoters and polynucleotides are the subject of this application. The novel genes are transformed into the relevant species using standard procedures. Their effects are measured by standard assays for the specific chemical/biochemical products.

These polynucleotides and proteins of the invention that participate in the relevant pathways and are useful for changing production of the above chemicals and biochemicals are

identified in the Reference tables by their enzyme function. More specifically, proteins of the invention that have the following enzymatic activities are of interest to modulate the corresponding pathways to produce precursors or final products noted above that are of industrial use. Biological activities of particular interest are listed below.

Other polynucleotides and proteins that regulate where, when and to what extent a pathway is active in a plant are extremely useful for modulating the synthesis and accumulation of valuable chemicals. These elements including transcription factors, proteins involved in signal transduction and other proteins in the control of gene expression are described elsewhere in this application.

Pathway Name	Enzyme Description	Comments
Alkaloid biosynthesis I	Morphine 6-dehydrogenase	Also acts on other alkaloids, including codeine, normorphine and ethylmorphine, but only very slowly on 7,8-saturated derivatives such as dihydromorphine and dihydrocodeine In the reverse direction, also reduces naloxone to the 6-alpha- hydroxy analog Activated by 2-mercaptoethanol
	Codeinone reductase (NADPH)	Stereospecifically catalyses the reversible reduction of codeinone to codeine, which is a direct precursor of morphine in the opium poppy plant, Papaver somniferum
	Salutaridine reductase (NADPH)	Stereospecifically catalyses the reversible reduction of salutaridine to salutaridinol, which is a direct precursor of morphinan alkaloids in the poppy plant, Papaver somniferum
	(S)-stylopine synthase	Catalyses an oxidative reaction that does not incorporate oxygen into the product Forms the second methylenedioxy bridge of the protoberberine alkaloid stylopine from oxidative ring closure of adjacent phenolic and methoxy groups of cheilanthifoline
	(S)-cheilanthifoline synthase	Catalyses an oxidative reaction that does not incorporate oxygen into the product Forms the methylenedioxy bridge of the protoberberine alkaloid cheilanthifoline from oxidative ring closure of adjacent phenolic and methoxy groups of scoulerine
	Salutaridine synthase	Forms the morphinan alkaloid salutaridine by intramolecular phenol oxidation of reticuline without the incorporation of oxygen into the product

	(S)-canadine synthase	Catalyses an oxidative reaction that does not incorporate oxygen into the product Oxidation of the methoxyphenol group of the alkaloid tetrahydrocolumbamine results in the formation of the methylenedioxy bridge of canadine
	Protopine 6-monooxygenase	Involved in benzophenanthridine alkaloid synthesis in higher plants
	Dihydrosanguinarine 10-monooxygenase	Involved in benzophenanthridine alkaloid synthesis in higher plants
	Monophenol monooxygenase	A group of copper proteins that also catalyse the reaction of EC 1.10.3.1, if only 1,2-benzenediols are available as substrate
	L-amino acid oxidase	
	1,2-dehydroreticulinium reductase (NADPH)	Stereospecifically reduces the 1,2-dehydroreticulinium ion to (R)-reticuline, which is a direct precursor of morphinan alkaloids in the poppy plant, papaver somniferum The enzyme does not catalyse the reverse reaction to any significant extent under physiological conditions
	Dihydrobenzophenanthridine oxidase	Also catalyzes: dihydrochelirubine + O(2) = chelirubine + H(2)O(2) Also catalyzes: dihydromacarpine + O(2) = macarpine + H(2)O(2) Found in higher plants Produces oxidized forms of the benzophenanthridine alkaloids
	Reticuline oxidase	The product of the reaction, (S)-scoulerine, is a precursor of protopine, protoberberine and benzophenanthridine alkaloid biosynthesis in plants Acts on (S)-reticuline and related compounds, converting the N- methyl group into the methylene bridge (`berberine bridge[PRIME]) of (S)-tetrahydroprotoberberines

	3[PRIME]-hydroxy-N-methyl-(S)-coclaurine 4[PRIME]-O-methyltransferase	Involved in isoquinoline alkaloid metabolism in plants Has also been shown to catalyse the methylation of (R,S)- laudanosoline, (S)-3[PRIME]-hydroxycoclaurine and (R,S)-7-O-methylnoraudanosoline
	(S)-scoulerine 9-O-methyltransferase	The product of this reaction is a precursor for protoberberine alkaloids in plants
	Columbamine O-methyltransferase	The product of this reaction is a protoberberine alkaloid that is widely distributed in the plant kingdom Distinct in specificity from EC 2.1.1.88
	10-hydroxydihydrosanguinarine 10-O-methyltransferase	Part of the pathway for synthesis of benzophenanthridine alkaloids in plants
	12-hydroxydihydrochelirubine 12-O-methyltransferase	Part of the pathway for synthesis of benzophenanthridine alkaloid macarpine in plants
	(R,S)-norcoclaurine 6-O-methyltransferase	Norcoclaurine is 6,7-dihydroxy-1-[(4-hydroxyphenyl)methyl]-1,2,3,4-tetrahydroisoquinoline The enzyme will also catalyse the 6-O-methylation of (R,S)- norlaudanosoline to form 6-O-methyl-norlaudanosoline, but this alkaloid has not been found to occur in plants
	Salutaridinol 7-O-acetyltransferase	At higher pH values the product, 7-O-acetylsalutaridinol, spontaneously closes the 4->5 oxide bridge by allylic elimination to form the morphine precursor thebaine From the opium poppy plant, Papaver somniferum
	Aspartate aminotransferase	Also acts on L-tyrosine, L-phenylalanine and L-tryptophan. This activity can be formed from EC 2.6.1.57 by controlled proteolysis
	Tyrosine aminotransferase	L-phenylalanine can act instead of L-tyrosine The mitochondrial enzyme may be identical with EC 2.6.1.1 The three isoenzymic forms are interconverted by EC 3.4.22.4

	Aromatic amino acid transferase	L-methionine can also act as donor, more slowly Oxaloacetate can act as acceptor Controlled proteolysis converts the enzyme to EC 2.6.1.1
	Tyrosine decarboxylase	The bacterial enzyme also acts on 3-hydroxytyrosine and, more slowly, on 3-hydroxyphenylalanine
	Aromatic-L-amino-acid decarboxylase	Also acts on L-tryptophan, 5-hydroxy-L-tryptophan and dihydroxy- L-phenylalanine (DOPA)
Alkaloid biosynthesis II	Tropine dehydrogenase	Oxidizes other tropan-3-alpha-ols, but not the corresponding beta-derivatives
	Tropinone reductase	
	Hyoscyamine (6S)-dioxygenase	
	6-beta-hydroxyhyoscyamine epoxidase	
	Amine oxidase (copper-containing)	A group of enzymes including those oxidizing primary amines, diamines and histamine One form of EC 1.3.1.15 from rat kidney also catalyses this reaction
	Putrescine N-methyltransferase	
	Ornithine decarboxylase	
	Oxalyl-CoA decarboxylase	
	Phenylalanine ammonia-lyase	May also act on L-tyrosine
Androgen and estrogen metabolism	3-beta-hydroxy-delta(5)-steroid dehydrogenase	Acts on 3-beta-hydroxyandrost-5-en-17-one to form androst-4-ene- 3,17-dione and on 3-beta-hydroxypregn-5-en-20-one to form progesterone
	11-beta-hydroxysteroid dehydrogenase	
	Estradiol 17-alpha-dehydrogenase	
	3-alpha-hydroxy-5-beta-androstane-17-one 3-alpha-	

	dehydrogenase	
	3-alpha (17-beta)-hydroxysteroid dehydrogenase (NAD+)	Also acts on other 17-beta-hydroxysteroids, on the 3-alpha-hydroxy group of pregnanes and bile acids, and on benzene dihydrodiol Different from EC 1.1.1.50 or EC 1.1.1.213
	3-alpha-hydroxysteroid dehydrogenase (B-specific)	Acts on other 3-alpha-hydroxysteroids and on 9-, 11- and 15-hydroxyprostaglandin B-specific with respect to NAD(+) or NADP(+) (cf. EC 1.1.1.213)
	3(or 17)beta-hydroxysteroid dehydrogenase	Also acts on other 3-beta- or 17-beta-hydroxysteroids (cf EC 1.1.1.209)
	Estradiol 17 beta-dehydrogenase	Also acts on (S)-20-hydroxypregn-4-en-3-one and related compounds, oxidizing the (S)-20-group B-specific with respect to NAD(P)(+)
	Testosterone 17-beta-dehydrogenase	
	Testosterone 17-beta-dehydrogenase (NADP+)	Also oxidizes 3-hydroxyhexobarbital to 3-oxohexobarbital
	Steroid 11-beta-monooxygenase	Also hydroxylates steroids at the 18-position, and converts 18-hydroxycorticosterone into aldosterone
	Estradiol 6-beta-monooxygenase	
	Androst-4-ene-3,17-dione monooxygenase	Has a wide specificity A single enzyme from Cylindrocarpon radicicola (EC 1.14.13.54) catalyses both this reaction and that catalysed by EC 1.14.99.4
	3-oxo-5-alpha-steroid 4-dehydrogenase	
	3-oxo-5-beta-steroid 4-	

	dehydrogenase	
	UDP-glucuronosyltransferase	Family of enzymes accepting a wide range of substrates, including phenols, alcohols, amines and fatty acids Some of the activities catalysed were previously listed separately as EC 2.4.1.42, EC 2.4.1.59, EC 2.4.1.61, EC 2.4.1.76, EC 2.4.1.77, EC 2.4.1.84, EC 2.4.1.107 and EC 2.4.1.108 A temporary nomenclature for the various forms whose delineation is in a state of flux
	Steroid sulfotransferase	Broad specificity resembling EC 2.8.2.2, but also acts on estrone
	Alcohol sulfotransferase	Primary and secondary alcohols, including aliphatic alcohols, ascorbate, chloramphenicol, ephedrine and hydroxysteroids, but not phenolic steroids, can act as acceptors (cf. EC 2.8.2.15)
	Estrone sulfotransferase	
	Arylsulfatase	A group of enzymes with rather similar specificities
	Steryl-sulfatase	Also acts on some related steryl sulfates
	17-alpha-hydroxyprogesterone aldolase	
	Steroid delta-isomerase	
C21-Steroid hormone metabolism	3-beta-hydroxy-delta(5)-steroid dehydrogenase	Acts on 3-beta-hydroxyandrost-5-en-17-one to form androst-4-ene- 3,17-dione and on 3-beta-hydroxypregn-5-en-20-one to form progesterone
	11-beta-hydroxysteroid dehydrogenase	
	20-alpha-hydroxysteroid dehydrogenase	A-specific with respect to NAD(P)(+)
	3-alpha-hydroxysteroid dehydrogenase (B-	Acts on other 3-alpha-hydroxysteroids and on 9-, 11- and 15-hydroxyprostaglandin B-specific with

	specific)	respect to NAD(+) or NADP(+) (cf. EC 1.1.1.213)
	3-alpha(or 20-beta)-hydroxysteroid dehydrogenase	The 3-alpha-hydroxyl group or 20-beta-hydroxyl group of pregnane and androstane steroids can act as donors
	Steroid 11-beta-monooxygenase	Also hydroxylates steroids at the 18-position, and converts 18-hydroxycorticosterone into aldosterone
	Corticosterone 18-monooxygenase	
	Cholesterol monooxygenase (side-chain cleaving)	The reaction proceeds in three stages, with hydroxylation at C-20 and C-22 preceding scission of the side-chain at C-20
	Steroid 21-monooxygenase	
	Progesterone 11-alpha-monooxygenase	
	Steroid 17-alpha-monooxygenase	
	Cholestenone 5-beta-reductase	
	Cortisone beta-reductase	
	Progesterone 5-alpha-reductase	Testosterone and 20-alpha-hydroxy-4-pregnen-3-one can act in place of progesterone
	3-oxo-5-beta-steroid 4-dehydrogenase	
	Steroid delta-isomerase	
Flavonoids, stilbene and lignin biosynthesis	Coniferyl-alcohol dehydrogenase	Specific for coniferyl alcohol; does not act on cinnamyl alcohol, 4-coumaryl alcohol or sinapyl alcohol
	Cinnamyl-alcohol dehydrogenase	Acts on coniferyl alcohol, sinapyl alcohol, 4-coumaryl alcohol and

		cinnamyl alcohol (cf. EC 1.1.1.194)
	Dihydrokaempferol 4-reductase	Also acts, in the reverse direction, on (+)-dihydroquercetin and (+)-dihydromyricetin Each dihydroflavonol is reduced to the corresponding cis-flavon- 3,4-diol NAD(+) can act instead of NADP(+), more slowly Involved in the biosynthesis of anthocyanidins in plants
	Flavonone 4-reductase	Involved in the biosynthesis of 3-deoxyanthocyanidins from flavonones such as naringenin or eriodictyol
	Peroxidase	
	Caffeate 3,4-dioxygenase	
	Naringenin 3-dioxygenase	
	Trans-cinnamate 4-monooxygenase	Also acts on NADH, more slowly
	Trans-cinnamate 2-monooxygenase	
	Flavonoid 3[PRIME]-monooxygenase	Acts on a number of flavonoids, including naringenin and dihydrokaempferol Does not act on 4-coumarate or 4-coumaroyl-CoA
	Monophenol monooxygenase	A group of copper proteins that also catalyse the reaction of EC 1.10.3.1, if only 1,2-benzenediols are available as substrate
	Cinnamoyl-CoA reductase	Also acts on a number of substituted cinnamoyl esters of coenzyme A
	Caffeoyl-CoA O-methyltransferase	
	Luteolin O-methyltransferase	Also acts on luteolin-7-O-beta-D-glucoside
	Caffeate O-methyltransferase	3,4-dihydroxybenzaldehyde and catechol can act as acceptor, more

		slowly
	Apigenin 4[PRIME]-O-methyltransferase	Converts apigenin into acacetin Naringenin (5,7,4[PRIME]-trihydroxyflavonone) can also act as acceptor, more slowly
	Quercetin 3-O-methyltransferase	Specific for quercetin. Related enzymes bring about the 3-O-methylation of other flavonols, such as galangin and kaempferol
	Isoflavone-7-O-beta-glucoside 6[PRIME][PRIME]-O-malonyltransferase	The 6-position of the glucose residue of formononetin can also act as acceptor Some other 7-O-glucosides of isoflavones, flavones and flavonols can also act, more slowly
	Pinosylvin synthase	Not identical with EC 2.3.1.74 or EC 2.3.1.95
	Naringenin-chalcone synthase	In the presence of NADH and a reductase, 6[PRIME]-deoxychalcone is produced
	Trihydroxystilbene synthase	Not identical with EC 2.3.1.74 or EC 2.3.1.146
	Quinate O-hydroxycinnamoyltransferase	Caffeoyl-CoA and 4-coumaroyl-CoA can also act as donors, more slowly Involved in the biosynthesis of chlorogenic acid in sweet potato and, with EC 2.3.1.98 in the formation of caffeoyl-CoA in tomato
	Coniferyl-alcohol glucosyltransferase	Sinapyl alcohol can also act as acceptor
	2-coumarate O-beta-glucosyltransferase	Coumarinate (cis-2-hydroxycinnamate) does not act as acceptor
	Scopoletin glucosyltransferase	
	Flavonol-3-O-glucoside L-rhamnosyltransferase	Converts flavonol 3-O-glucosides to 3-O-rutinosides Also acts, more slowly, on rutin, quercetin 3-O-galactoside and flavonol O3-rhamnosides
	Flavone 7-O-beta-glucosyltransferase	A number of flavones, flavonones and flavonols can function as acceptors Different from EC 2.4.1.91
	Flavonol 3-O-glucosyltransferase	Acts on a variety of flavonols, including quercetin and quercetin 7-O-

		glucoside Different from EC 2.4.1.81
	Flavone apiosyltransferase	7-O-beta-D-glucosides of a number of flavonoids and of 4-substituted phenols can act as acceptors
	Coniferin beta-glucosidase	Also hydrolyzes syringin, 4-cinnamyl alcohol beta-glucoside, and, more slowly, some other aryl beta-glycosides A plant cell-wall enzyme involved in the biosynthesis of lignin
	Beta-glucosidase	Wide specificity for beta-D-glucosides. Some examples also hydrolyse one or more of the following: beta-D-galactosides, alpha-L- arabinosides, beta-D-xylosides, and beta-D-fucosides
	Chalcone isomerase	
	4-coumarate--CoA ligase	

Pathway Name	Enzyme Description	Enzyme Comments
Ascorbate and aldarate metabolism	D-threo-aldose 1-dehydrogenase	Acts on L-fucose, D-arabinose and L-xylose The animal enzyme was also shown to act on L-arabinose, and the enzyme from Pseudomonas caryophylli on L-glucose
	L-threonate 3-dehydrogenase	
	Glucuronate reductase	Also reduces D-galacturonate May be identical with EC 1.1.1.2
	Glucuronolactone reductase	
	L-arabinose 1-dehydrogenase	
	L-galactonolactone oxidase	Acts on the 1,4-lactones of L-galactonic, D-altronic, L-fuconic, D-arabinic and D-threonic acids Not identical with EC 1.1.3.8 (cf. EC 1.3.2.3)
	L-gulonolactone oxidase	The product spontaneously isomerizes to L-ascorbate
	L-ascorbate oxidase	
	L-ascorbate peroxidase	
	Ascorbate 2,3-dioxygenase	
	2,5-dioxovalerate dehydrogenase	

	Aldehyde dehydrogenase (NAD+)	Wide specificity, including oxidation of D-glucuronolactone to D-glucarate
	Galactonolactone dehydrogenase	Cf. EC 1.1.3.24
	Monodehydroascorbate reductase (NADH)	
	Glutathione dehydrogenase (ascorbate)	
	L-arabinonolactonase	
	Gluconolactonase	Acts on a wide range of hexono-1,5-lactones
	Uronolactonase	
	1,4-lactonase	Specific for 1,4-lactones with 4-8 carbon atoms Does not hydrolyse simple aliphatic esters, acetylcholine, sugar lactones or substituted aliphatic lactones, e.g. 3-hydroxy-4-butyrolactone
	2-dehydro-3-deoxyglucarate aldolase	
	L-arabinonate dehydratase	
	Glucarate dehydratase	
	5-dehydro-4-deoxyglucarate dehydratase	
	Galactarate dehydratase	
	2-dehydro-3-	

	deoxy-L-arabinonate dehydratase	
Carbon fixation	Malate dehydrogenase	Also oxidizes some other 2-hydroxydicarboxylic acids
	Malate dehydrogenase (decarboxylating)	Does not decarboxylates added oxaloacetate
	Malate dehydrogenase (oxaloacetate decarboxylating) (NADP+)	Also decarboxylates added oxaloacetate
	Malate dehydrogenase (NADP+)	Activated by light
	Glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating)	
	Transketolase	Wide specificity for both reactants, e.g. converts hydroxypyruvate and R-CHO into CO(2) and R-CHOH-CO-CH(2)OH Transketolase from Alcaligenes faecalis shows high activity with D-erythrose as acceptor
	Aspartate aminotransferase	Also acts on L-tyrosine, L-phenylalanine and L-tryptophan. This activity can be formed from EC 2.6.1.57 by controlled

		proteolysis
	Alanine aminotransferase	2-aminobutanoate acts slowly instead of alanine
	Sedoheptulokinase	
	Phosphoribulokinase	
	Pyruvate kinase	UTP, GTP, CTP, ITP and dATP can also act as donors Also phosphorylates hydroxylamine and fluoride in the presence of CO(2)
	Phosphoglycerate kinase	
	Pyruvate,phosphate dikinase	
	Fructose-bisphosphatase	The animal enzyme also acts on sedoheptulose 1,7-bisphosphate
	Sedoheptulose-bisphosphatase	
	Phosphoenolpyruvate carboxylase	
	Ribulose-bisphosphate carboxylase	Will utilize O(2) instead of CO(2), forming 3-phospho-D-glycerate and 2-phosphoglycolate
	Phosphoenolpyruvate carboxykinase (ATP)	
	Fructose-bisphosphate aldolase	Also acts on (3S,4R)-ketose 1-phosphates The yeast and bacterial enzymes are zinc proteins The enzymes increase electron-attraction by the carbonyl

		group, some (Class I) forming a protonated imine with it, others (Class II), mainly of microbial origin, polarizing it with a metal ion, e.g zinc
	Phosphoketolase	
	Ribulose-phosphate 3-epimerase	Also converts D-erythrose 4-phosphate into D-erythrulose 4-phosphate and D-threose 4-phosphate
	Triosephosphate isomerase	
	Ribose 5-phosphate epimerase	Also acts on D-ribose 5-diphosphate and D-ribose 5-triphosphate
Phenylalanine metabolism	(R)-4-hydroxyphenyllactate dehydrogenase	Also acts, more slowly, on (R)-3-phenyllactate, (R)-3-(indole-3- yl)lactate and (R)-lactate
	Hydroxyphenylpyruvate reductase	Also acts on 3-(3,4-dihydroxyphenyl)lactate Involved with EC 2.3.1.140 in the biosynthesis of rosmarinic acid
	Aryl-alcohol dehydrogenase	A group of enzymes with broad specificity towards primary alcohols with an aromatic or cyclohex-1-ene ring, but with low or no activity towards short-chain aliphatic alcohols
	Peroxidase	
	Catechol 1,2-	Involved in the

	dioxygenase	metabolism of nitro-aromatic compounds by a strain of Pseudomonas putida
	2,3-dihydroxybenzoate 3,4-dioxygenase	
	3-carboxyethylcatechol 2,3-dioxygenase	
	Catechol 2,3-dioxygenase	The enzyme from Alcaligines sp. strain O-1 has also been shown to catalyse the reaction: 3-Sulfocatechol + O(2) + H(2)O = 2-hydroxymuconate + bisulfite. It has been referred to as 3-sulfocatechol 2,3-dioxygenase. Further work will be necessary to show whether or not this is a distinct enzyme
	4-hydroxyphenylpyruvate dioxygenase	
	Protocatechuate 3,4-dioxygenase	
	Hydroxyquinol 1,2-dioxygenase	The product isomerizes to 2-maleylacetate (cis-hex-2-enedioate) Highly specific; catechol and pyrogallol are acted on at less than 1% of the rate at which benzene-1,2,4-triol

		is oxidized
	Protocatechuate 4,5-dioxygenase	
	Phenylalanine 2-monooxygenase	Also catalyses a reaction similar to that of EC 1.4.3.2, forming 3-phenylpyruvate, NH(3) and H(2)O(2), but more slowly
	Anthranilate 1,2-dioxygenase (deaminating, decarboxylating)	
	Benzoate 1,2-dioxygenase	A system, containing a reductase which is an iron-sulfur flavoprotein (FAD), and an iron-sulfur oxygenase
	Toluene dioxygenase	A system, containing a reductase which is an iron-sulfur flavoprotein (FAD), an iron-sulfur oxygenase, and a ferredoxin Some other aromatic compounds, including ethylbenzene, 4-xylene and some halogenated toluenes, are converted into the corresponding cis-dihydrodiols
	Naphthalene 1,2-dioxygenase	A system, containing a reductase which is an iron-sulfur flavoprotein (FAD), an iron-sulfur oxygenase, and ferredoxin
	Benzene 1,2-dioxygenase	A system, containing a reductase which is

		an iron-sulfur flavoprotein, an iron-sulfur oxygenase and ferredoxin
	Salicylate 1-monooxygenase	
	Trans-cinnamate 4-monooxygenase	Also acts on NADH, more slowly
	Benzoate 4-monooxygenase	
	4-hydroxybenzoate 3-monooxygenase	Most enzymes from Pseudomonas are highly specific for NAD(P)H (cf EC 1.14.13.33)
	3-hydroxybenzoate 4-monooxygenase	Also acts on a number of analogs of 3-hydroxybenzoate substituted in the 2, 4, 5 and 6 positions
	3-hydroxybenzoate 6-monooxygenase	Also acts on a number of analogs of 3-hydroxybenzoate substituted in the 2, 4, 5 and 6 positions NADPH can act instead of NADH, more slowly
	4-hydroxybenzoate 3-monooxygenase (NAD(P)H)	The enzyme from Corynebacterium cyclohexanicum is highly specific for 4-hydroxybenzoate, but uses NADH and NADPH at approximately equal rates (cf. EC 1.14.13.2). It is less specific for NADPH than EC 1.14.13.2
	Anthranilate 3-monooxygenase (deaminating)	The enzyme from Aspergillus niger is an iron protein; that

		from the yeast Trichosporon cutaneum is a flavoprotein (FAD)
	Melilotate 3-monooxygenase	
	Phenol 2-monooxygenase	Also active with resorcinol and O-cresol
	Mandelate 4-monooxygenase	
	3-hydroxybenzoate 2-monooxygenase	
	4-cresol dehydrogenase (hydroxylating)	Phenazine methosulfate can act as acceptor A quinone methide is probably formed as intermediate The product is oxidized further to 4-hydroxybenzoate
	Benzaldehyde dehydrogenase (NAD+)	
	Aminomuconate-semialdehyde dehydrogenase	Also acts on 2-hydroxymuconate semialdehyde
	Phenylacetaldehyde dehydrogenase	
	4-carboxy-2-hydroxymuconate-6-semialdehyde dehydrogenase	Does not act on unsubstituted aliphatic or aromatic aldehydes or glucose NAD(+) can replace NADP(+), but with lower affinity
	Aldehyde dehydrogenase (NAD(P)+)	
	Benzaldehyde dehydrogenase	

	(NADP+)	
	Coumarate reductase	
	Cis-1,2-dihydrobenzene-1,2-diol dehydrogenase	
	Cis-1,2-dihydro-1,2-dihydroxynaphthalene dehydrogenase	Also acts, at half the rate, on cis-anthracene dihydrodiol and cis-phenanthrene dihydrodiol
	2-enoate reductase	Acts, in the reverse direction, on a wide range of alkyl and aryl alpha,beta-unsaturated carboxylate ions 2-butenoate was the best substrate tested
	Maleylacetate reductase	
	Phenylalanine dehydrogenase	The enzyme from Bacillus badius and Sporosarcina ureae are highly specific for L-phenylalanine, that from Bacillus sphaericus also acts on L-tyrosine
	L-amino acid oxidase	
	Amine oxidase (flavin-containing)	Acts on primary amines, and usually also on secondary and tertiary amines
	Amine oxidase (copper-containing)	A group of enzymes including those oxidizing primary amines, diamines and histamine One form of EC 1.3.1.15 from rat kidney also catalyses this

		reaction
	D-amino-acid dehydrogenase	Acts to some extent on all D-amino acids except D-aspartate and D-glutamate
	Aralkylamine dehydrogenase	Phenazine methosulfate can act as acceptor Acts on aromatic amines and, more slowly, on some long-chain aliphatic amines, but not on methylamine or ethylamine (cf EC 1.4.99.3)
	Glutamine N-phenylacetyltransferase	
	Acetyl-CoA C-acyltransferase	
	D-amino-acid N-acetyltransferase	
	Phenylalanine N-acetyltransferase	Also acts, more slowly, on L-histidine and L-alanine
	Glycine N-benzoyltransferase	Not identical with EC 2.3.1.13 or EC 2.3.1.68
	Aspartate aminotransferase	Also acts on L-tyrosine, L-phenylalanine and L-tryptophan. This activity can be formed from EC 2.6.1.57 by controlled proteolysis
	D-alanine aminotransferase	Acts on the D-isomers of leucine, aspartate, glutamate, aminobutyrate, norvaline and asparagine

	Tyrosine aminotransferase	L-phenylalanine can act instead of L-tyrosine The mitochondrial enzyme may be identical with EC 2.6.1.1 The three isoenzymic forms are interconverted by EC 3.4.22.4
	Aromatic amino acid transferase	L-methionine can also act as donor, more slowly Oxaloacetate can act as acceptor Controlled proteolysis converts the enzyme to EC 2.6.1.1
	Histidinol-phosphate aminotransferase	
	3-oxoadipate CoA-transferase	
	3-oxoadipate enol-lactonase	Acts on the product of EC 4.1.1.44
	Carboxymethylenebutenolidase	
	2-pyrone-4,6-dicarboxylate lactonase	The product isomerizes to 4-oxalmesaconate
	Hippurate hydrolase	Acts on various N-benzoylamino acids
	Amidase	
	Acylphosphatase	
	2-hydroxymuconate-semialdehyde hydrolase	
	Aromatic-L-amino-acid decarboxylase	Also acts on L-tryptophan, 5-hydroxy-L-tryptophan and dihydroxy- L-

		phenylalanine (DOPA)
	Phenylpyruvate decarboxylase	Also acts on indole-3-pyruvate
	4-carboxymuconolactone decarboxylase	
	O-pyrocatechuate decarboxylase	
	Phenylalanine decarboxylase	Also acts on tyrosine and other aromatic amino acids
	4-hydroxybenzoate decarboxylase	
	Protocatechuate decarboxylase	
	Benzoylformate decarboxylase	
	4-oxalocrotonate decarboxylase	Involved in the meta-cleavage pathway for the degradation of phenols, cresols and catechols
	4-hydroxy-4-methyl-2-oxoglutarate aldolase	Also acts on 4-hydroxy-4-methyl-2-oxoadipate and 4-carboxy-4-hydroxy-2-oxohexadioate
	2-oxopent-4-enoate hydratase	Also acts, more slowly, on cis-2-oxohex-4-enoate, but not on the trans-isomer
	Phenylalanine ammonia-lyase	May also act on L-tyrosine
	Phenylalanine racemase (ATP-hydrolysing)	
	Mandelate racemase	
	Phenylpyruvate	Also acts on other

	tautomerase	arylpyruvates
	5-carboxymethyl-2-hydroxymuconate delta-isomerase	
	Muconolactone delta-isomerase	
	Muconate cycloisomerase	Also acts, in the reverse reaction, on 3-methyl-cis-cis-hexa- dienedioate and, very slowly, on cis-trans-hexadienedioate Not identical with EC 5.5.1.7 or EC 5.5.1.11
	3-carboxy-cis,cis-muconate cycloisomerase	
	Carboxy-cis,cis-muconate cyclase	
	Chloromuconate cycloisomerase	Spontaneous elimination of HCl produces cis-4-carboxymethylenebut-2- en-4-olide Also acts in reverse direction on 2-chloro-cis,cis-muconate Not identical with EC 5.5.1.1 or EC 5.5.1.11
	Phenylacetate--CoA ligase	Phenoxyacetate can replace phenylacetate
	Benzoate--CoA ligase	Also acts on 2-, 3- and 4-fluorobenzoate, but only very slowly on the corresponding chlorobenzoates

	4-hydroxybenzoate--CoA ligase	
	Phenylacetate--CoA ligase	Also acts, more slowly, on acetate, propanoate and butanoate, but not on hydroxy derivatives of phenylacetate and related compounds
Phenylalanine, tyrosine and tryptophan biosynthesis	Quinate 5-dehydrogenase	
	Shikimate 5-dehydrogenase	
	Quinate dehydrogenase (pyrroloquinoline-quinone)	
	Phenylalanine 4-monooxygenase	
	Prephenate dehydrogenase	This enzyme in the enteric bacteria also possesses chorismate mutase activity (EC 5.4.99.5) and converts chorismate into prephenate
	Prephenate dehydrogenase (NADP+)	
	Cyclohexadienyl dehydrogenase	Also acts on prephenate and D-prephenyllactate (cf. EC 1.3.1.12)
	2-methyl-branched-chain-enoyl-CoA reductase	From Ascaris suum The reaction proceeds only in the presence of another flavoprotein (ETF=[PRIME]Electron-Transferring Flavoprotein[PRIME])
	Phenylalanine	The enzyme from

	dehydrogenase	Bacillus badius and Sporosarcina ureae are highly specific for L-phenylalanine, that from Bacillus sphaericus also acts on L-tyrosine
	L-amino acid oxidase	
	Anthranilate phosphoribosyltransferase	In some organisms, this enzyme is part of a multifunctional protein together with one or more components of the system for biosynthesis of tryptophan (EC 4.1.1.48, EC 4.1.3.27, EC 4.2.1.20, and EC 5.3.1.24)
	3-phosphoshikimate 1-carboxyvinyltransferase	
	Aspartate aminotransferase	Also acts on L-tyrosine, L-phenylalanine and L-tryptophan. This activity can be formed from EC 2.6.1.57 by controlled proteolysis
	Tyrosine aminotransferase	L-phenylalanine can act instead of L-tyrosine The mitochondrial enzyme may be identical with EC 2.6.1.1 The three isoenzymic forms are interconverted by EC 3.4.22.4

	Aromatic amino acid transferase	L-methionine can also act as donor, more slowly Oxaloacetate can act as acceptor Controlled proteolysis converts the enzyme to EC 2.6.1.1
	Histidinol-phosphate aminotransferase	
	Shikimate kinase	
	Indole-3-glycerol-phosphate synthase	In some organisms, this enzyme is part of a multifunctional protein together with one or more components of the system for biosynthesis of tryptophan (EC 2.4.2.18, EC 4.1.3.27, EC 4.2.1.20, and EC 5.3.1.24)
	2-dehydro-3-deoxyphosphoheptonate aldolase	
	Anthranilate synthase	In some organisms, this enzyme is part of a multifunctional protein together with one or more components of the system for biosynthesis of tryptophan (EC 2.4.2.18, EC 4.1.1.48, EC 4.2.1.20, and EC 5.3.1.24) The native enzyme in the complex with uses

		either glutamine or (less efficiently) NH(3). The enzyme separated from the complex uses NH(3) only
	3-dehydroquinate dehydratase	
	Phosphopyruvate hydratase	Also acts on 3-phospho-D-erythronate
	Tryptophan synthase	Also catalyses the conversion of serine and indole into tryptophan and water and of indoleglycerol phosphate into indole and glyceraldehyde phosphate In some organisms, this enzyme is part of a multifunctional protein together with one or more components of the system for biosynthesis of tryptophan (EC 2.4.2.18, EC 4.1.1.48, EC 4.1.3.27, and EC 5.3.1.24)
	Prephenate dehydratase	This enzyme in the enteric bacteria also possesses chorismate mutase activity and converts chorismate into prephenate
	Carboxycyclohexadienyl dehydratase	Also acts on prephenate and D-prephenyllactate Cf. EC 4.2.1.51

	3-dehydroquinate synthase	The hydrogen atoms on C-7 of the substrate are retained on C-2 of the products
	Chorismate synthase	Shikimate is numbered so that the double-bond is between C-1 and C-2, but some earlier papers numbered in the reverse direction
	Phosphoribosylanthranilate isomerase	In some organisms, this enzyme is part of a multifunctional protein together with one or more components of the system for biosynthesis of tryptophan (EC 2.4.2.18, EC 4.1.1.48, EC 4.1.3.27, and EC 4.2.1.20)
	Chorismate mutase	
	Tyrosine--tRNA ligase	
	Phenylalanine--tRNA ligase	
Starch and sucrose metabolism	UDP-glucose 6-dehydrogenase	Also acts on UDP-2-deoxyglucose
	Glucoside 3-dehydrogenase	The enzyme acts on D-glucose, D-galactose, D-glucosides and D-galactosides, but D-glucosides react more rapidly than D-galactosides
	CDP-4-dehydro-6-deoxyglucose reductase	Two proteins are involved but no partial reaction has been observed in the presence of

		either alone
	Phosphorylase	The recommended name should be qualified in each instance by adding the name of the natural substance, e.g. maltodextrin phosphorylase, starch phosphorylase, glycogen phosphorylase
	Levansucrase	Some other sugars can act as D-fructosyl acceptors
	Glycogen (starch) synthase	The recommended name varies according to the source of the enzyme and the nature of its synthetic product Glycogen synthase from animal tissues is a complex of a catalytic subunit and the protein glycogenin The enzyme requires glucosylated glycogenin as a primer; this is the reaction product of EC 2.4.1.186 A similar enzyme utilizes ADP-glucose (Cf. EC 2.4.1.21)
	Cellulose synthase (UDP-forming)	Involved in the synthesis of cellulose A similar enzyme utilizes GDP-glucose (Cf. EC 2.4.1.29)
	Sucrose synthase	

	Sucrose-phosphate synthase	
	Alpha,alpha-trehalose-phosphate synthase (UDP-forming)	See also EC 2.4.1.36
	UDP-glucuronosyltransferase	Family of enzymes accepting a wide range of substrates, including phenols, alcohols, amines and fatty acids Some of the activities catalysed were previously listed separately as EC 2.4.1.42, EC 2.4.1.59, EC 2.4.1.61, EC 2.4.1.76, EC 2.4.1.77, EC 2.4.1.84, EC 2.4.1.107 and EC 2.4.1.108 A temporary nomenclature for the various forms whose delineation is in a state of flux
	1,4-alpha-glucan branching enzyme	Converts amylose into amylopectin The recommended name requires a qualification depending on the product, glycogen or amylopectin, e.g. glycogen branching enzyme, amylopectin branching enzyme. The latter has frequently been termed Q-enzyme

	Cellobiose phosphorylase	
	Starch (bacterial glycogen) synthase	The recommended name various according to the source of the enzyme and the nature of its synthetic product, e.g. starch synthase, bacterial glycogen synthase A similar enzyme utilizes UDP-glucose (Cf. EC 2.4.1.11)
	4-alpha-glucanotransferase	An enzymic activity of this nature forms part of the mammalian and Yeast glycogen branching system (see EC 3.2.1.33)
	Cellulose synthase (GDP-forming)	Involved in the synthesis of cellulose A similar enzyme utilizes UDP-glucose (Cf. EC 2.4.1.12)
	1,3-beta-glucan synthase	
	Phenol beta-glucosyltransferase	Acts on a wide range of phenols
	Amylosucrase	
	Polygalacturonate 4-alpha-galacturonosyltransferase	
	Dextransucrase	
	Alpha,alpha-trehalose phosphorylase	
	Sucrose phosphorylase	In the forward reaction, arsenate may replace phosphate In the

		reverse reaction various ketoses and L-arabinose may replace D-fructose
	Maltose phosphorylase	
	1,4-beta-D-xylan synthase	
	Hexokinase	D-glucose, D-mannose, D-fructose, sorbitol and D-glucosamine can act as acceptors ITP and dATP can act as donors The liver isoenzyme has sometimes been called glucokinase
	Phosphoglucokinase	
	Glucose-1,6-bisphosphate synthase	D-glucose 6-phosphate can act as acceptor, forming D-glucose 1,6-bisphosphate
	Glucokinase	A group of enzymes found in invertebrates and microorganisms highly specific for glucose
	Fructokinase	
	Glucose-1-phosphate phosphodismutase	
	Protein-N(PI)-phosphohistidine-sugar phosphotransferase	Comprises a group of related enzymes The protein substrate is a phosphocarrier protein of low molecular mass (9.5 Kd) A phosphoenzyme intermediate is

		formed The enzyme translocates the sugar it phosphorylates into bacteria Aldohexoses and their glycosides and alditols are phosphorylated on O-6; fructose and sorbose on O-1 Glycerol and disaccharides are also substrates
	Glucose-1-phosphate adenylyltransferase	
	Glucose-1-phosphate cytidylyltransferase	
	Glucose-1-phosphate guanylyltransferase	Also acts, more slowly, on D-mannose 1-phosphate
	UTP--glucose-1-phosphate uridylyltransferase	
	Pectinesterase	
	Trehalose-phosphatase	
	Sucrose-phosphatase	
	Glucose-6-phosphatase	Wide distribution in animal tissues Also catalyses potent transphosphorylations from carbamoyl phosphate, hexose phosphates, pyrophosphate, phosphoenolpyruvate and nucleoside di- and triphosphates,

		to D-glucose, D-mannose, 3-methyl-D-glucose, or 2-deoxy-D-glucose (cf. EC 2.7.1.62, EC 2.7.1.79, and EC 3.9.1.1)
	Alpha-amylase	Acts on starch, glycogen and related polysaccharides and oligosaccharides in a random manner; reducing groups are liberated in the alpha-configuration
	Oligo-1,6-glucosidase	Also hydrolyses palatinose The enzyme from intestinal mucosa is a single polypeptide chain also catalysing the reaction of EC 3.2.1.48
	Maltose-6[PRIME]-phosphate glucosidase	Hydrolyses a variety of 6-phospho-D-glucosides, including maltose 6-phosphate, alpha[PRIME]alpha-trehalose 6-phosphate, sucrose 6-phosphate and p-nitrophenyl-alpha-D-glucopyranoside 6-phosphate (as a chromogenic substrate) The enzyme is activated by Fe(II), Mn(II), Co(II) and Ni(II). It is rapidly inactivated in air
	Polygalacturonase	
	Beta-amylase	Acts on starch,

		glycogen and related polysaccharides and oligosaccharides producing beta-maltose by an inversion
	Alpha-glucosidase	Group of enzymes whose specificity is directed mainly towards the exohydrolysis of 1,4-alpha-glucosidic linkages, and that hydrolyse oligosaccharides rapidly, relative to polysaccharides, which are hydrolysed relatively slowly, or not at all The intestinal enzyme also hydrolyses polysaccharides, catalysing the reactions of EC 3.2.1.3, and, more slowly, hydrolyses 1,6-alpha- D-glucose links
	Beta-glucosidase	Wide specificity for beta-D-glucosides. Some examples also hydrolyse one or more of the following: beta-D-galactosides, alpha-L- arabinosides, beta-D-xylosides, and beta-D-fucosides
	Beta-fructofuranosidase	Substrates include sucrose Also catalyses fructotransferase

		reactions
	Alpha,alpha-trehalase	
	Glucan 1,4-alpha-glucosidase	Most forms of the enzyme can rapidly hydrolyse 1,6-alpha-D-glucosidic bonds when the next bond in sequence is 1,4, and some preparations of this enzyme hydrolyse 1,6- and 1,3-alpha-D-glucosidic bonds in other polysaccharides This entry covers all such enzymes acting on polysaccharides more rapidly than on oligosaccharides EC 3.2.1.20 from mammalian intestine can catalyse similar reactions
	Beta-glucuronidase	
	Amylo-1,6-glucosidase	In mammals and yeast this enzyme is linked to a glycosyltransferase similar to EC 2.4.1.25; together these two activities constitute the glycogen debranching system
	Xylan 1,4-beta-xylosidase	Also hydrolyses xylobiose Some other exoglycosidase activities have been found associated with this enzyme in sheep liver

	Glucan endo-1,3-beta-D-glucosidase	Very limited action on mixed-link (1,3-1,4-)-beta-D-glucans Hydrolyses laminarin, paramylon and pachyman Different from EC 3.2.1.6
	Cellulase	Will also hydrolyse 1,4-linkages in beta-D-glucans also containing 1,3-linkages
	Sucrose alpha-glucosidase	This enzyme is isolated from intestinal mucosa as a single polypeptide chain also displaying activity towards isomaltose (oligo-1,6-glucosidase, cf. EC 3.2.1.10)
	Cyclomaltodextrinase	Also hydrolyses linear maltodextrin
	Glucan 1,3-beta-glucosidase	Acts on oligosaccharides but very slowly on laminaribiose
	Levanase	
	Galacturan 1,4-alpha-galacturonidase	
	Glucan 1,4-beta-glucosidase	Acts on 1,4-beta-D-glucans and related oligosaccharides Cellobiose is hydrolysed, very slowly
	Cellulose 1,4-beta-cellobiosidase	
	Alpha,alpha-phosphotrehalase	
	ADP-sugar diphosphatase	Has a distinct specificity from the

		UDP-sugar pyrophosphatase (EC 3.6.1.45)
	Nucleotide pyrophosphatase	Substrates include NAD(+), NADP(+), FAD, CoA and also ATP and ADP
	UDP-glucuronate decarboxylase	
	CDP-glucose 4,6-dehydratase	
	CDP-abequose epimerase	
	UDP-glucuronate 4-epimerase	
	Glucose-6-phosphate isomerase	Also catalyses the anomerization of D-glucose 6-phosphate
	Phosphoglucomutase	Maximum activity is only obtained in the presence of alpha-D-glucose 1,6-bisphosphate. This bisphosphate is an intermediate in the reaction, being formed by transfer of a phosphate residue from the enzyme to the substrate, but the dissociation of bisphosphate from the enzyme complex is much slower than the overall isomerization Also, more slowly, catalyses the interconversion of 1-phosphate and 6-phosphate isomers of many other alpha-D-hexoses, and the

		interconversion of alpha-D-ribose 1-phosphate and 5-phosphate
	Beta-phosphoglucomutase	
	Maltose alpha-D-glucosyltransferase	
Tryptophan metabolism	Indole-3-lactate dehydrogenase	
	Indole-3-acetaldehyde reductase (NADH)	
	Indole-3-acetaldehyde reductase (NADPH)	
	3-hydroxyacyl-CoA dehydrogenase	Also oxidizes S-3-hydroxyacyl-N-acylthioethanolamine and S-3-hydroxyacylhydrolipoate Some enzymes act, more slowly, with NADP(+) Broad specificity to acyl chain-length (cf. EC 1.1.1.211)
	O-aminophenol oxidase	Isophenoxazine may be formed by a secondary condensation from the initial oxidation product
	Catalase	This enzyme can also act as a peroxidase (EC 1.11.1.7) for which several organic substances, especially ethanol, can act as a

		hydrogen donor A manganese protein containing Mn(III) in the resting state, which also belongs here, is often called pseudocatalase Enzymes from some microorganisms, such as Penicillium simplicissimum, which exhibit both catalase and peroxidase activity, have sometimes been referred to as catalase-peroxidase
	7,8-dihydroxykynurenate 8,8A-dioxygenase	
	Tryptophan 2,3-dioxygenase	Broad specificity towards tryptamine and derivatives including D- and L-tryptophan, 5-hydroxytryptophan and serotonin
	Indole 2,3-dioxygenase	The enzyme from jasminum is a flavoprotein containing copper, and forms anthranilate as the final product One enzyme from Tecoma stans is also a flavoprotein containing copper and uses three atoms of oxygen per molecule of indole, to form anthranil (3,4-benzisoxazole) A second enzyme from Tecoma stans,

		which is not a flavoprotein, uses four atoms of oxygen and forms anthranilate as the final product
	2,3-dihydroxyindole 2,3-dioxygenase	
	Indoleamine-pyrrole 2,3-dioxygenase	Acts on many substituted and unsubstituted indoleamines, including melatonin Involved in the degradation of melatonin
	3-hydroxyanthranilate 3,4-dioxygenase	The product of the reaction spontaneously rearrange to quinolinic acid (quin)
	Tryptophan 2-monooxygenase	
	Tryptophan 2[PRIME]-dioxygenase	Acts on a number of indolyl-3-alkane derivatives, oxidizing the 3-side-chain in the 2[PRIME]-position. Best substrates are L-tryptophan and 5-hydroxy-L-tryptophan
	Kynurenine 3-monooxygenase	
	Unspecific monooxygenase	Acts on a wide range of substrates including many xenobiotics, steroids, fatty acids, vitamins and prostaglandins Reactions catalysed include hydroxylation,

		epoxidation, N-oxidation, sulfooxidation, N-, S- and O-dealkylations, desulfation, deamination, and reduction of azo, nitro, and N-oxide groups
	Anthranilate 3-monooxygenase	
	Tryptophan 5-monooxygenase	Activated by phosphorylation, catalysed by a CA(2+)-activated protein kinase
	Kynurenine 7,8-hydroxylase	
	Aldehyde dehydrogenase (NAD+)	Wide specificity, including oxidation of D-glucuronolactone to D-glucarate
	Aminomuconate-semialdehyde dehydrogenase	Also acts on 2-hydroxymuconate semialdehyde
	Aldehyde oxidase	Also oxidizes quinoline and pyridine derivatives May be identical with EC 1.1.3.22
	Indole-3-acetaldehyde oxidase	Also oxidizes indole-3-aldehyde and acetaldehyde, more slowly
	Oxoglutarate dehydrogenase (lipoamide)	Component of the multienzyme 2-oxoglutarate dehydrogenase complex
	Kynurenate-7,8-dihydrodiol dehydrogenase	
	Glutaryl-CoA dehydrogenase	

	L-amino acid oxidase	
	Amine oxidase (flavin-containing)	Acts on primary amines, and usually also on secondary and tertiary amines
	Amine oxidase (copper-containing)	A group of enzymes including those oxidizing primary amines, diamines and histamine One form of EC 1.3.1.15 from rat kidney also catalyses this reaction
	Acetylindoxyl oxidase	
	Acetylserotonin O-methyltransferase	Some other hydroxyindoles also act as acceptor, more slowly
	Indole-3-pyruvate C-methyltransferase	
	Amine N-methyltransferase	A wide range of primary, secondary, and tertiary amines can act as acceptors, including tryptamine, aniline, nicotine and a variety of drugs and other xenobiotics
	Aralkylamine N-acetyltransferase	Narrow specificity towards aralkylamines, including serotonin Not identical with EC 2.3.1.5
	Acetyl-CoA C-acetyltransferase	
	Tryptophan aminotransferase	Also acts on 5-hydroxytryptophan and, to a lesser

		extent on the phenyl amino acids
	Kynurenine--oxoglutarate aminotransferase	Also acts on 3-hydroxykynurenine
	Thioglucosidase	Has a wide specificity for thioglycosides
	Amidase	
	Formamidase	Also acts, more slowly, on acetamide, propanamide and butanamide
	Arylformamidase	Also acts on other aromatic formylamines
	Nitrilase	Acts on a wide range of aromatic nitriles including (indole-3- yl)-acetonitrile and also on some aliphatic nitriles, and on the corresponding acid amides (cf. EC 4.2.1.84)
	Kynureninase	Also acts on 3[PRIME]-hydroxykynurenine and some other (3-arylcarbonyl)-alanines
	Aromatic-L-amino-acid decarboxylase	Also acts on L-tryptophan, 5-hydroxy-L-tryptophan and dihydroxy- L-phenylalanine (DOPA)
	Phenylpyruvate decarboxylase	Also acts on indole-3-pyruvate
	Aminocarboxymuconate-semialdehyde	The product rearranges non-enzymically to

	decarboxylase	picolinate
	Tryptophanase	Also catalyses the synthesis of tryptophan from indole and serine Also catalyses 2,3-elimination and beta-replacement reactions of some indole-substituted tryptophan analogs of L-cysteine, L-serine and other 3-substituted amino acids
	Enoyl-CoA hydratase	Acts in the reverse direction With cis-compounds, yields (3R)-3-hydroxyacyl-CoA (cf. EC 4.2.1.74)
	Nitrile hydratase	Acts on short-chain aliphatic nitriles, converting them into the corresponding acid amides Does not act on these amides or on aromatic nitriles (cf EC 3.5.5.1)
	Tryptophan--tRNA ligase	
Tyrosine metabolism	Alcohol dehydrogenase	Acts on primary or secondary alcohols or hemiacetals The animal, but not the yeast, enzyme acts also on cyclic secondary alcohols
	(R)-4-hydroxyphenyllactate dehydrogenase	Also acts, more slowly, on (R)-3-phenyllactate, (R)-3-(indole-3- yl)lactate and (R)-lactate
	Hydroxyphenylpyruvate	Also acts on 3-(3,4-dihydroxyphenyl)lact

	reductase	ate Involved with EC 2.3.1.140 in the biosynthesis of rosmarinic acid
	Aryl-alcohol dehydrogenase	A group of enzymes with broad specificity towards primary alcohols with an aromatic or cyclohex-1-ene ring, but with low or no activity towards short-chain aliphatic alcohols
	Catechol oxidase	Also acts on a variety of substituted catechols Many of these enzymes also catalyse the reaction listed under EC 1.14.18.1; this is especially true for the classical tyrosinase
	Iodide peroxidase	
	3,4-dihydroxyphenyl acetate 2,3-dioxygenase	
	4-hydroxyphenylpyruvate dioxygenase	
	Stizolobate synthase	The intermediate product undergoes ring closure and oxidation, with NAD(P)(+) as acceptor, to stizolobic acid
	Stizolobinate synthase	The intermediate product undergoes ring closure and oxidation, with NAD(P)(+) as

		acceptor, to stizolobinic acid
	Gentisate 1,2-dioxygenase	
	Homogentisate 1,2-dioxygenase	
	4-hydroxyphenylacetate 1-monooxygenase	Also acts on 4-hydroxyhydratropate forming 2-methylhomogentisate and on 4-hydroxyphenoxyacetate forming hydroquinone and glycolate
	4-hydroxyphenylacetate 3-monooxygenase	
	Tyrosine N-monooxygenase	
	Hydroxyphenylacetonitrile 2-monooxygenase	
	Tyrosine 3-monooxygenase	Activated by phosphorylation, catalysed by EC 2.7.1.128
	Dopamine-beta-monooxygenase	Stimulated by fumarate
	Monophenol monooxygenase	A group of copper proteins that also catalyse the reaction of EC 1.10.3.1, if only 1,2-benzenediols are available as substrate
	Succinate-semialdehyde dehydrogenase (NAD(P)+)	
	Aryl-aldehyde dehydrogenase	Oxidizes a number of aromatic aldehydes, but not aliphatic aldehydes

	Aldehyde dehydrogenase (NAD+)	Wide specificity, including oxidation of D-glucuronolactone to D-glucarate
	4-carboxy-2-hydroxymuconate-6-semialdehyde dehydrogenase	Does not act on unsubstituted aliphatic or aromatic aldehydes or glucose NAD(+) can replace NADP(+), but with lower affinity
	Aldehyde dehydrogenase (NAD(P)+)	
	4-hydroxyphenylacetaldehyde dehydrogenase	With EC 4.2.1.87, brings about the metabolism of octopamine in Pseudomonas
	Aldehyde oxidase	Also oxidizes quinoline and pyridine derivatives May be identical with EC 1.1.3.22
	L-amino acid oxidase	
	Amine oxidase (flavin-containing)	Acts on primary amines, and usually also on secondary and tertiary amines
	Amine oxidase (copper-containing)	A group of enzymes including those oxidizing primary amines, diamines and histamine One form of EC 1.3.1.15 from rat kidney also catalyses this reaction
	Aralkylamine dehydrogenase	Phenazine methosulfate can act as acceptor Acts on aromatic amines and, more slowly, on

		some long-chain aliphatic amines, but not on methylamine or ethylamine (cf EC 1.4.99.3)
	Phenol O-methyltransferase	Acts on a wide variety of simple alkyl-, methoxy- and halo-phenols
	Tyramine N-methyltransferase	Has some activity on phenylethylamine analogs
	Phenylethanolamine N-methyltransferase	Acts on various phenylethanolamines; converts noradrenalin into adrenalin
	Catechol O-methyltransferase	The mammalian enzymes act more rapidly on catecholamines such as adrenaline or noradrenaline than on catechols
	Glutamine N-phenylacetyltransferase	
	Rosmarinate synthase	Involved with EC 1.1.1.237 in the biosynthesis of rosmarinic acid
	Hydroxymandelonitrile glucosyltransferase	3,4-dihydroxymandelonitrile can also act as acceptor
	Aspartate aminotransferase	Also acts on L-tyrosine, L-phenylalanine and L-tryptophan. This activity can be formed from EC 2.6.1.57 by controlled proteolysis
	Dihydroxyphenylalanine	

	aminotransferase	
	Tyrosine aminotransferase	L-phenylalanine can act instead of L-tyrosine The mitochondrial enzyme may be identical with EC 2.6.1.1 The three isoenzymic forms are interconverted by EC 3.4.22.4
	Aromatic amino acid transferase	L-methionine can also act as donor, more slowly Oxaloacetate can act as acceptor Controlled proteolysis converts the enzyme to EC 2.6.1.1
	Histidinol-phosphate aminotransferase	
	Fumarylacetoacetase	Also acts on other 3,5- and 2,4-dioxo acids
	Acylpyruvate hydrolase	Acts on formylpyruvate, 2,4-dioxopentanoate, 2,4-dioxohexanoate and 2,4-dioxoheptanoate
	Tyrosine decarboxylase	The bacterial enzyme also acts on 3-hydroxytyrosine and, more slowly, on 3-hydroxyphenylalanine
	Aromatic-L-amino-acid decarboxylase	Also acts on L-tryptophan, 5-hydroxy-L-tryptophan and dihydroxy- L-

		phenylalanine (DOPA)
	Gentisate decarboxylase	
	5-oxopent-3-ene-1,2,5-tricarboxylate decarboxylase	
	Tyrosine phenol-lyase	Also slowly catalyses pyruvate formation from D-tyrosine, S-methyl-L-cysteine, L-cysteine, L-serine and D-serine
	(S)-norcoclaurine synthase	The reaction makes a 6-membered ring by forming a bond between C-6 of the 3,4-dihydroxyphenyl group of the dopamine and C-1 of the aldehyde in the imine formed between the substrates The product is the precursor of the benzylisoquinoline alkaloids in plants Will also catalyse the reaction of 4-(2-aminoethyl)benzene-1,2-diol + (3,4-dihydroxyphenyl)acetaldehyde to form (S)-norlaudanosoline, but this alkaloid has not been found to occur in plants
	Dihydroxyphenylalanine ammonia-lyase	
	Phenylalanine ammonia-lyase	May also act on L-tyrosine

	Maleylacetoacetate isomerase	Also acts on maleylpyruvate
	Maleylpyruvate isomerase	
	Phenylpyruvate tautomerase	Also acts on other arylpyruvates
	5-carboxymethyl-2-hydroxymuconate delta-isomerase	
	Tyrosine 2,3-aminomutase	
	Phenylacetate--CoA ligase	Also acts, more slowly, on acetate, propanoate and butanoate, but not on hydroxy derivatives of phenylacetate and related compounds

IX. PROMOTERS AS SENTINELS

Useful promoters include those that are capable of facilitating preferential transcription, i.e. tissue-specific or developmentally regulated gene expression and being a component of facile systems to evaluate the metabolic/physiological state of a plant cell, tissue or organ. Many such promoters are included in this application. Operably linking a sequence to these promoters that can act as a reporter and inserting the construct into a plant allows detection of the preferential in plantar transcription. For example, the quantitative state of responses to environmental conditions can be detected by using a plant having a construct that contains a stress-inducible promoter linked to and controlling expression of a sequence encoding GFP. The greater the stress promoter is induced, the greater the levels of fluorescence from GFP will be produced and this provides a measure of the level of stress being expressed by the plant and/or the ability of the plant to respond internally to the stress.

More specifically, using this system the activities of any metabolic pathway (catabolic and anabolic), stress-related pathways as on any plant gene repeated activity can be monitored. In addition, assays can be developed using this sentinel system to select for superior genotypes with greater yield characteristics or to select for plants with altered responses to chemical, herbicide, or plant growth regulators or to identify chemical, herbicides or plant growth regulators by their response on such sentinels.

Specifically, a promoter that is regulated in plants in the desired way, is operably linked to a reporter such as GFP, RFP, etc., and the constructs are introduced into the plant of interest. The behavior of the reporter is monitored using technologies typically specific for that reporter. With GFP, RFP, etc., it could typically be by microscopy of whole plants, organs, tissues or cells under excitation by an appropriate wavelength of UV light.

BRIEF DESCRIPTION OF THE TABLES

BRIEF DESCRIPTION OF THE TABLES

1. Reference and Sequence Tables

The sequences of exemplary SDFs and polypeptides corresponding to the coding sequences of the instant invention are described in the Reference and Sequence Tables (sometimes referred to as the REF and SEQ Tables. The Reference Table refers to a number of "Maximum Length Sequences" or "MLS." Each MLS corresponds to the longest cDNA obtained, either by cloning or by the prediction from genomic sequence. The sequence of the MLS is the cDNA sequence as described in the Av subsection of the Reference Table.

The Reference Table includes the following information relating to each MLS:

I. cDNA Sequence
- A. 5' UTR
- B. Coding Sequence
- C. 3' UTR

II. Genomic Sequence
- A. Exons
- B. Introns
- C. Promoters

III. Link of cDNA Sequences to Clone IDs

IV. Multiple Transcription Start Sites

V. Polypeptide Sequences
- A. Signal Peptide
- B. Domains
- C. Related Polypeptides

VI. Related Polynucleotide Sequences

I. cDNA SEQUENCE

The Reference Table indicates which sequence in the Sequence Table represents the sequence of each MLS. The MLS sequence can comprise 5' and 3' UTR as well as coding sequences. In addition, specific cDNA clone numbers also are included in the Reference Table when the MLS sequence relates to a specific cDNA clone.

A. 5' UTR

The location of the 5' UTR can be determined by comparing the most 5' MLS sequence with the corresponding genomic sequence as indicated in the Reference Table. The sequence that matches, beginning at any of the transcriptional start sites and ending at the last nucleotide before any of the translational start sites corresponds to the 5' UTR.

B. Coding Region

The coding region is the sequence in any open reading frame found in the MLS. Coding regions of interest are indicated in the PolyP SEQ subsection of the Reference Table.

C. 3' UTR

The location of the 3' UTR can be determined by comparing the most 3' MLS sequence with the corresponding genomic sequence as indicated in the Reference Table. The sequence that matches, beginning at the translational stop site and ending at the last nucleotide of the MLS corresponds to the 3' UTR.

II. GENOMIC SEQUENCE

Further, the Reference Table indicates the specific "gi" number of the genomic sequence if the sequence resides in a public databank. For each genomic sequence, Reference tables indicate which regions are included in the MLS. These regions can include the 5' and 3' UTRs as well as the coding sequence of the MLS. See, for example, the scheme below:

```
                Region 1                 Region 2          Region 3
             ________________          __________       ________________
---------| 5' UTR    | Exon |---------|   Exon    |--------| Exon | 3' UTR |-------
             ________________          __________       ________________
    ^                          ^                     ^                         ^
    |                 ^        |                     |           ^             |
Promoter              |     Intron                Intron         |
               Translational                               Stop Codon
                Start Site
```

The Reference Table reports the first and last base of each region that are included in an MLS sequence. An example is shown below:

gi No. 47000:

37102 … 37497

37593 … 37925

The numbers indicate that the MLS contains the following sequences from two regions of gi No. 47000; a first region including bases 37102-37497, and a second region including bases 37593-37925.

A. EXON SEQUENCES

The location of the exons can be determined by comparing the sequence of the regions from the genomic sequences with the corresponding MLS sequence as indicated by the Reference Table.

i. INITIAL EXON

To determine the location of the initial exon, information from the

(1) polypeptide sequence section;

(2) cDNA polynucleotide section; and

(3) the genomic sequence section

of the Reference Table is used. First, the polypeptide section will indicate where the translational start site is located in the MLS sequence. The MLS sequence can be matched to the genomic sequence that corresponds to the MLS. Based on the match between the MLS and corresponding genomic sequences , the location of the translational start site can be determined in one of the regions of the genomic sequence. The location of this translational start site is the start of the first exon.

Generally, the last base of the exon of the corresponding genomic region, in which the translational start site was located, will represent the end of the initial exon. In some cases, the initial exon will end with a stop codon, when the initial exon is the only exon.

In the case when sequences representing the MLS are in the positive strand of the corresponding genomic sequence, the last base will be a larger number than the first base. When the sequences representing the MLS are in the negative strand of the corresponding genomic sequence, then the last base will be a smaller number than the first base.

ii. INTERNAL EXONS

Except for the regions that comprise the 5' and 3' UTRs, initial exon, and terminal exon, the remaining genomic regions that match the MLS sequence are the internal exons. Specifically, the bases defining the boundaries of the remaining regions also define the intron/exon junctions of the internal exons.

iii. TERMINAL EXON

As with the initial exon, the location of the terminal exon is determined with information from the

(1) polypeptide sequence section;

(2) cDNA polynucleotide section; and

(3) the genomic sequence section

of the Reference Table. The polypeptide section will indicate where the stop codon is located in the MLS sequence. The MLS sequence can be matched to the corresponding genomic sequence. Based on the match between MLS and corresponding genomic sequences, the location of the stop codon can be determined in one of the regions of the genomic sequence. The location of this stop codon is the end of the terminal exon. Generally, the first base of the exon of the corresponding genomic region that matches the cDNA sequence, in which the stop codon was located, will represent the beginning of the terminal exon. In some cases, the translational start site will represent the start of the terminal exon, which will be the only exon.

In the case when the MLS sequences are in the positive strand of the corresponding genomic sequence, the last base will be a larger number than the first base. When the MLS sequences are in the negative strand of the corresponding genomic sequence, then the last base will be a smaller number than the first base.

B. INTRON SEQUENCES

In addition, the introns corresponding to the MLS are defined by identifying the genomic sequence located between the regions where the genomic sequence comprises exons. Thus, introns are defined as starting one base downstream of a genomic region comprising an exon, and end one base upstream from a genomic region comprising an exon.

C. PROMOTER SEQUENCES

As indicated below, promoter sequences corresponding to the MLS are defined as sequences upstream of the first exon; more usually, as sequences upstream of the first of multiple transcription start sites; even more usually as sequences about 2,000 nucleotides upstream of the first of multiple transcription start sites.

III. LINK of cDNA SEQUENCES to CLONE IDs

As noted above, the Reference Table identifies the cDNA clone(s) that relate to each MLS. The MLS sequence can be longer than the sequences included in the cDNA clones. In such a case, the Reference Table indicates the region of the MLS that is included in the clone. If either the 5' or 3' termini of the cDNA clone sequence is the same as the MLS sequence, no mention will be made.

IV. Multiple Transcription Start Sites

Initiation of transcription can occur at a number of sites of the gene. The Reference Table indicates the possible multiple transcription sites for each gene. In the Reference Table, the location of the transcription start sites can be either a positive or negative number.

The positions indicated by positive numbers refer to the transcription start sites as located in the MLS sequence. The negative numbers indicate the transcription start site within the genomic sequence that corresponds to the MLS.

To determine the location of the transcription start sites with the negative numbers, the MLS sequence is aligned with the corresponding genomic sequence. In the instances when a public genomic sequence is referenced, the relevant corresponding genomic sequence can be found by direct reference to the nucleotide sequence indicated by the "gi" number shown in the public genomic DNA section of the Reference Table. When the position is a negative number, the transcription start site is located in the corresponding genomic sequence upstream of the base that matches the beginning of the MLS sequence in the alignment. The negative number is relative to the first base of the MLS sequence which matches the genomic sequence corresponding to the relevant "gi" number.

In the instances when no public genomic DNA is referenced, the relevant nucleotide sequence for alignment is the nucleotide sequence associated with the amino acid sequence designated by "gi" number of the later PolyP SEQ subsection.

V. Polypeptide Sequences

The PolyP SEQ subsection lists SEQ ID NOs and Ceres SEQ ID NO for polypeptide sequences corresponding to the coding sequence of the MLS sequence and the location of the translational start site with the coding sequence of the MLS sequence.

The MLS sequence can have multiple translational start sites and can be capable of producing more than one polypeptide sequence.

A. Signal Peptide

The Reference tables also indicate in subsection (B) the cleavage site of the putative signal peptide of the polypeptide corresponding to the coding sequence of the MLS sequence. Typically, signal peptide coding sequences comprise a sequence encoding the first residue of the polypeptide to the cleavage site residue.

B. Domains

Subsection (C) provides information regarding identified domains (where present) within the polypeptide and (where present) a name for the polypeptide domain.

C. Related Polypeptides

Subsection (Dp) provides (where present) information concerning amino acid sequences that are found to be related and have some percentage of sequence identity to the polypeptide sequences of the Reference and Sequence Tables. These related sequences are identified by a "gi" number.

VI. Related Polynucleotide Sequences

Subsection (Dn) provides polynucleotide sequences (where present) that are related to and have some percentage of sequence identity to the MLS or corresponding genomic sequence.

Abbreviation	Description
Max Len. Seq.	Maximum Length Sequence
rel to	Related to
Clone Ids	Clone ID numbers
Pub gDNA	Public Genomic DNA
gi No.	gi number
Gen. Seq. in Cdna	Genomic Sequence in cDNA (Each region for a single gene prediction is listed on a separate line. In the case of multiple gene predictions, the group of regions relating to a single prediction are separated by a blank line)
(Ac) cDNA SEQ	cDNA sequence

Abbreviation	Description
- Pat. Appln. SEQ ID NO	Patent Application SEQ ID NO:
- Ceres SEQ ID NO: 1673877	Ceres SEQ ID NO:
- SEQ # w. TSS	Location within the cDNA sequence, SEQ ID NO:, of Transcription Start Sites which are listed below
- Clone ID #: # -> #	Clone ID comprises bases # to # of the cDNA Sequence
PolyP SEQ	Polypeptide Sequence
- Pat. Appln. SEQ ID NO:	Patent Application SEQ ID NO:
- Ceres SEQ ID NO	Ceres SEQ ID NO:
- Loc. SEQ ID NO: @ nt.	Location of translational start site in cDNA of SEQ ID NO: at nucleotide number
(C) Pred. PP Nom. & Annot.	Nomination and Annotation of Domains within Predicted Polypeptide(s)
- (Title)	Name of Domain
- Loc. SEQ ID NO #: # -> # aa.	Location of the domain within the polypeptide of SEQ ID NO: from # to # amino acid residues.
(Dp) Rel. AA SEQ	Related Amino Acid Sequences
- Align. NO	Alignment number
- gi No	Gi number
- Desp.	Description
- % Idnt.	Percent identity
- Align. Len.	Alignment Length
- Loc. SEQ ID NO: # -> # aa	Location within SEQ ID NO: from # to # amino acid residue.

2. Protein Group Table

This table indicates groups of proteins that share a signature sequence (also referred to as a consensus sequence). The Protein group also referred to as the Ortholog group is named by the peptide ID with which all members were compared. Each group contains sequences that were included at the 10^{-50}, 10^{-30}, and 10^{-10} p-value cutoffs. For each group, the peptide ID and at which cutoff the peptide was included into the group. The same peptide ID may be included in the group three times as peptide ID 50, peptide ID 30 and peptide ID 10. The data indicates that peptide ID was included in the group when the threshold was either 10^{-50}, 10^{-30}, or 10^{-10}. All the peptide IDs that are followed by "50" were included in the protein group when the e-value cutoff was 10^{-50}. All the peptide IDs that are followed by either "30" or "50" were included in the protein group when the threshold e-value was 10^{-30}. All the peptide IDs that are followed by "10", "30" or "50" were included in the protein group when 10^{-10} was used as the e-value cutoff.

At the end of each protein group is a list of the consensus sequence that proteins share at the 10^{-50}, 10^{-30}, or 10^{-10}. The consensus sequence contains both lower-case and upper-case letters. The upper-case letters represent the standard one-letter amino acid abbreviations. The lower case letters represent classes of amino acids:

- "t" refers to tiny amino acids, which are specifically alanine, glycine, serine and threonine.
- "p"refers to polar amino acids, which are specifically, asparagine and glutamine
- "n" refers to negatively charged amino acids, which are specifically, aspartic acid and glutamic acid
- "+" refers to positively charged residues, which are specifically, lysine, arginine, and histidine
- "r" refers to aromatic residues, which are specifically, phenylalanine, tyrosine, and tryptophan,
- "a" refers to aliphatic residues, which are specifically, isoleucine, valine, leucine, and methonine

3. Protein Group_Matrix Table

In addition to each consensus sequence, Applicants have generated a scoring matrix to provide further description of the consensus sequence. The first row of each matrix indicates the residue position in the consensus sequence. The matrix reports number of occurrences of all the amino acids that were found in the group members for every residue position of the signature sequence. The matrix also indicates for each residue position, how many different organisms were found to have a polypeptide in the group that included a residue at the relevant position.

The last line of the matrix indicates all the amino acids that were found at each position of the consensus.

4. MA_diff Table

The MA_diff Table presents the results of the differential expression experiments for the mRNAs, as reported by their corresponding cDNA ID number, that were differentially transcribed under a particular set of conditions as compared to a control sample. The cDNA ID numbers correspond to those utilized in the Reference and Sequence Tables. Increases in mRNA abundance levels in experimental plants versus the controls are denoted with the plus sign (+). Likewise, reductions in mRNA abundance levels in the experimental plants are denoted with the minus (-) sign.

The Table is organized according to each set of experimental conditions, which are denoted by the term "Expt ID:" followed by a particular number. The table below links each Expt ID with a short description of the experiment and the parameters.

For each experiment ID a method of the normalization is specified. "Method: 2" represents normalization by median the goal of the method is to adjust the ratios by a factor so that the median of the ratio distribution is 1. Method 3 is the normalization procedure conducted by Aglilent Technologies, Inc. Palo Alto, California, USA.

The MA_diff Table also specifies the specific parameters and the experiment number (e.g. 107871) used in compiling the data. The experiment numbers are referenced in the appropriate utility/functions sections herein. The background threshold was set to "BKG_Threshold=X" to reduce the effect of the background on the signal.

Finally, the Table includes reference to an "Organism_ID" number. This number refers to the cDNA spotted on the chip were similar to Arabidopsis thaliana (3769) sequences or whether the oligo used for the chips were similar to Zea mays (311987)sequences.

5. MA_diff (Experiment) Tables

Example No.	Experiment short name	genome	Expt ID	Value	Parameter	Units

3ii	3642-1	Arabidopsis	108512	3746-1	Plant Line	Hours
3n	Arab_0.001%_MeJA_1	Arabidopsis	108568	Aerial	Tissue	Tissue
				0.001%_MeJA	Treatment	Compound
				1	Timepoint	Hours
3n	Arab_0.001%_MeJA_1	Arabidopsis	108569	Aerial	Tissue	Tissue
				6	Timepoint	Hours
				0.001%_MeJA	Treatment	Compound
3j	Arab_0.1uM_Epi-Brass_1	Arabidopsis	108580	Aerial	Tissue	Tissue
				1	Timepoint	Hours
				0.1uM_Brassino_Steroid	Treatment	Compound
3j	Arab_0.1uM_Epi-Brass_1	Arabidopsis	108581	Aerial	Tissue	Tissue
				6	Timepoint	Hours
				0.1uM_Brassino_Steroid	Treatment	Compound
3g	Arab_100uM_ABA_1	Arabidopsis	108560	Aerial	Tissue	Tissue
				1	Timepoint	Hours
				100uM_ABA	Treatment	Compound
3g	Arab_100uM_ABA_1	Arabidopsis	108561	Aerial	Tissue	Tissue
				100uM_ABA	Treatment	Compound
				6	Timepoint	Hours
3I	Arab_100uM_BA_1	Arabidopsis	108566	Aerial	Tissue	Tissue
				1	Timepoint	Hours
				100uM_BA	Treatment	Compound
3I	Arab_100uM_BA_1	Arabidopsis	108567	Aerial	Tissue	Tissue
				100uM_BA	Treatment	Compound
				6	Timepoint	Hours
3k	Arab_100uM_GA3_1	Arabidopsis	108562	Aerial	Tissue	Tissue
				1	Timepoint	Hours
				100uM GA3	Treatment	Compound
3k	Arab_100uM_GA3_1	Arabidopsis	108563	Aerial	Tissue	Tissue
				100uM GA3	Treatment	Compound
				6	Timepoint	Hours

3h	Arab_100uM_NAA_1	Arabidopsis	108564	Aerial	Tissue	Tissue
				1	Timepoint	Hours
				100uM_NAA	Treatment	Compound
3h	Arab_100uM_NAA_1	Arabidopsis	108565	Aerial	Tissue	Tissue
				100uM_NAA	Treatment	Compound
				6	Timepoint	Hours
3r	Arab_20%_PEG_1	Arabidopsis	108570	Aerial	Tissue	Tissue
				1	Timepoint	Hours
				20%PEG	Treatment	Compound
3r	Arab_20%_PEG_1	Arabidopsis	108571	Aerial	Tissue	Tissue
				20%PEG	Treatment	Compound
				6	Timepoint	Hours
3o	Arab_2mM_SA_1	Arabidopsis	108586	Aerial	Tissue	Tissue
				2mM_SA	Treatment	Compound
				1	Timepoint	Hours
3o	Arab_2mM_SA_1	Arabidopsis	108587	Aerial	Tissue	Tissue
				6	Timepoint	Hours
				2mM_SA	Treatment	Compound
3u	Arab_5mM_H2O2_1	Arabidopsis	108582	Aerial	Tissue	Tissue
				1	Timepoint	Hours
				5mM_H2O2	Treatment	Compound
3u	Arab_5mM_H2O2_1	Arabidopsis	108583	Aerial	Tissue	Tissue
				5mM_H2O2	Treatment	Compound
				6	Timepoint	Hours
3v	Arab_5mM_NaNP_1	Arabidopsis	108584	Aerial	Tissue	Tissue
				1	Timepoint	Hours
				5mM_NaNP	Treatment	Compound
3v	Arab_5mM_NaNP_1	Arabidopsis	108585	Aerial	Tissue	Tissue
				5mM_NaNP	Treatment	Compound
				6	Timepoint	Hours
3t	Arab_Cold_1	Arabidopsis	108578	Aerial	Tissue	Tissue
				Cold	Treatment	Compound
				1	Timepoint	Hours
3t	Arab_Cold_1	Arabidopsis	108579	Aerial	Tissue	Tissue

				6	Timepoint	Hours
				Cold	Treatment	Compound
3g	Arab_Drought_1	Arabidopsis	108572	Aerial	Tissue	Tissue
				1	Timepoint	Hours
				Drought	Treatment	Compound
3g	Arab_Drought_1	Arabidopsis	108573	Aerial	Tissue	Tissue
				Drought	Treatment	Compound
				6	Timepoint	Hours
3s	Arab_Heat_1	Arabidopsis	108576	Aerial	Tissue	Tissue
				1	Timepoint	Hours
				Heat (42 deg C)	Treatment	Compound
3s	Arab_Heat_1	Arabidopsis	108577	Aerial	Tissue	Tissue
				Heat (42 deg C)	Treatment	Compound
				6	Timepoint	Hours
3aa (ovule)	Arab_Ler-pi_ovule_1	Arabidopsis	108595	Ler_pi	Plant Line	Hours
				Ovule	Tissue	Tissue
3b	Arab_Ler-rhl_root_1	Arabidopsis	108594	Ler_rhl	Plant Line	Hours
				Root	Tissue	Tissue
31	Arab_NO3_H-to-L_1	Arabidopsis	108592	Aerial	Tissue	Tissue
				Low Nitrogen	Treatment	Compound
				12	Timepoint	Hours
31	Arab_NO3_H-to-L_1	Arabidopsis	108593	Aerial	Tissue	Tissue
				24	Timepoint	Hours
				Low Nitrogen	Treatment	Compound
31	Arab_NO3_L-to-H_1	Arabidopsis	108588	Aerial	Tissue	Tissue
				2	Timepoint	Hours
				Nitrogen	Treatment	Compound
31	Arab_NO3_L-to-H_1	Arabidopsis	108589	Aerial	Tissue	Tissue
				Nitrogen	Treatment	Compound
				6	Timepoint	Hours
31	Arab_NO3_L-to-H_1	Arabidopsis	108590	Aerial	Tissue	Tissue
				9	Timepoint	Hours
				Nitrogen	Treatment	Compound

3l	Arab_NO3_L-to-H_1	Arabidopsis	108591	Aerial	Tissue	Tissue
				Nitrogen	Treatment	Compound
				12	Timepoint	Hours
3p	Arab_Wounding_1	Arabidopsis	108574	Aerial	Tissue	Tissue
				1	Timepoint	Hours
				Wounding	Treatment	Compound
3p	Arab_Wounding_1	Arabidopsis	108575	Aerial	Tissue	Tissue
				Wounding	Treatment	Compound
				6	Timepoint	Hours
3o	Columbia/CS3726 flower SA	Arabidopsis	108475	Columbia	species	Hours
				SA	Treatment	Compound
				5 weeks	Timepoint	Hours
3o	Columbia/CS3726 flower SA	Arabidopsis	108476	CS3726	species	Hours
				5 weeks	Timepoint	Hours
				SA	Treatment	Compound
3p	Corn_0.001Percent_MeJA	Zea Mays	108555	Aerial	Tissue	Tissue
				24	Timepoint	Hours
				0.001%_MeJA	Treatment	Compound
3j	Corn_0.1uM_Brassino_Steroid	Zea Mays	108557	24	Timepoint	Hours
				Aerial	Tissue	Tissue
				0.1uM_Brassino_Steroid	Treatment	Compound
3g	Corn_100uM_ABA	Zea Mays	108513	Aerial	Tissue	Tissue
				ABA	Treatment	Compound
				6	Timepoint	Hours
3g	Corn_100uM_ABA	Zea Mays	108597	Aerial	Tissue	Tissue
				24	Timepoint	Hours
				100uM_ABA	Treatment	Compound
3i	Corn_100uM_BA	Zea Mays	108517	Aerial	Tissue	Tissue
				6	Timepoint	Hours
				BA	Treatment	Compound
3k	Corn_100uM_GA3	Zea Mays	108519	Aerial	Tissue	Tissue
				100uM	Treatment	Compound

				Giberillic Acid		
				1	Timepoint	Hours
3k	Corn_100uM_GA3	Zea Mays	108520	Aerial	Tissue	Tissue
				6	Timepoint	Hours
				100uM Giberillic Acid	Treatment	Compound
3k	Corn_100uM_GA3	Zea Mays	108521	Aerial	Tissue	Tissue
				100uM Giberillic Acid	Treatment	Compound
				12	Timepoint	Hours
3h	Corn_100uM_NAA	Zea Mays	108516	Aerial	Tissue	Tissue
				NAA	Treatment	Compound
				6	Timepoint	Hours
3h	Corn_100uM_NAA	Zea Mays	108554	Aerial	Tissue	Tissue
				24	Timepoint	Hours
				NAA	Treatment	Compound
3hh	Corn_1400-6/S-17	Zea Mays	108598	Shoot apices	Tissue	Tissue
3r	Corn_150mM_NaCl	Zea Mays	108541	Aerial	Tissue	Tissue
				1	Timepoint	Hours
				150mM_NaCl	Treatment	Compound
3r	Corn_150mM_NaCl	Zea Mays	108542	Aerial	Tissue	Tissue
				150mM_NaCl	Treatment	Compound
				6	Timepoint	Hours
3r	Corn_150mM_NaCl	Zea Mays	108553	Aerial	Tissue	Tissue
				24	Timepoint	Hours
				150mM_NaCl	Treatment	Compound
3r	Corn_20%_PEG	Zea Mays	108539	Aerial	Tissue	Tissue
				1	Timepoint	Hours
				20%PEG	Treatment	Compound
3r	Corn_20%_PEG	Zea Mays	108540	Aerial	Tissue	Tissue
				20%PEG	Treatment	Compound
				6	Timepoint	Hours

3o	Corn_2mM_SA	Zea Mays	108515	Aerial	Tissue	Tissue
				SA	Treatment	Compound
				12	Timepoint	Hours
3o	Corn_2mM_SA	Zea Mays	108552	Aerial	Tissue	Tissue
				SA	Treatment	Compound
				24	Timepoint	Hours
3u	Corn_5mM_H2O2	Zea Mays	108537	Aerial	Tissue	Tissue
				H2O2	Treatment	Compound
				1	Timepoint	Hours
3u	Corn_5mM_H2O2	Zea Mays	108538	Aerial	Tissue	Tissue
				6	Timepoint	Hours
				H2O2	Treatment	Compound
3u	Corn_5mM_H2O2	Zea Mays	108558	Aerial	Tissue	Tissue
				24	Timepoint	Hours
				H2O2	Treatment	Compound
3v	Corn_5mM_NO	Zea Mays	108526	Aerial	Tissue	Tissue
				NO	Treatment	Compound
				1	Timepoint	Hours
3v	Corn_5mM_NO	Zea Mays	108527	Aerial	Tissue	Tissue
				6	Timepoint	Hours
				NO	Treatment	Compound
3v	Corn_5mM_NO	Zea Mays	108559	Aerial	Tissue	Tissue
				12	Timepoint	Hours
				NO	Treatment	Compound
3t	Corn_Cold	Zea Mays	108533	Aerial	Tissue	Tissue
				1	Timepoint	Hours
				Cold	Treatment	Compound
3t	Corn_Cold	Zea Mays	108534	Aerial	Tissue	Tissue
				Cold	Treatment	Compound
				6	Timepoint	Hours
3q	Corn_Drought	Zea Mays	108502	Drought	Treatment	Compound
				1	Timepoint	Hours
3q	Corn_Drought	Zea Mays	108503	Drought	Treatment	Compound
				6	Timepoint	Hours
3q	Corn_Drought	Zea Mays	108504	Drought	Treatment	Compound
				12	Timepoint	Hours
3q	Corn_Drought	Zea Mays	108556	Drought	Treatment	Compound
				24	Timepoint	Hours

3s	Corn_Heat	Zea Mays	108522	Aerial	Tissue	Tissue
				1	Timepoint	Hours
				Heat (42 deg C)	Treatment	Compound
3s	Corn_Heat	Zea Mays	108523	Aerial	Tissue	Tissue
				6	Timepoint	Hours
				Heat (42 deg C)	Treatment	Compound
3gg	Corn_Imbibed Seeds	Zea Mays	108518	Imbibed	Treatment	Compound
				4	Age	days old
				Roots	Tissue	Tissue
3gg	Corn_Imbibed Seeds	Zea Mays	108528	Imbibed	Treatment	Compound
				Aerial	Tissue	Tissue
				5	Age	days old
3gg	Corn_Imbibed Seeds	Zea Mays	108529	Imbibed	Treatment	Compound
				5	Age	days old
				Root	Tissue	Tissue
3gg	Corn_Imbibed Seeds	Zea Mays	108530	Imbibed	Treatment	Compound
				Aerial	Tissue	Tissue
				6	Age	days old
3gg	Corn_Imbibed Seeds	Zea Mays	108531	Imbibed	Treatment	Compound
				6	Age	days old
				root	Tissue	Tissue
3gg	Corn_Imbibed Seeds	Zea Mays	108545	Imbibed	Treatment	Compound
				Aerial	Tissue	Tissue
				3	Age	days old
3gg	Corn_Imbibed Seeds	Zea Mays	108546	Imbibed	Treatment	Compound
				3	Age	days old
				Root	Tissue	Tissue
3gg	Corn_Imbibed Seeds	Zea Mays	108547	Imbibed	Treatment	Compound
				Aerial	Tissue	Tissue
				4	Age	days old
3gg	Corn_Imbibed_Embryo_Endosperm	Zea Mays	108543	2	Age	days old
				Imbibed	Treatment	Compound
				Embryo	Tissue	Tissue

3gg	Corn_Imbibed_Embryo_Endosperm	Zea Mays	108544	2	Age	days old
				Endosperm	Tissue	Tissue
				Imbibed	Treatment	Compound
3ee	Corn_Meristem	Zea Mays	108535	Root Meristem	Tissue	Tissue
				192	Timepoint	Hours
3ee	Corn_Meristem	Zea Mays	108536	Shoot Meristem	Tissue	Tissue
				192	Timepoint	Hours
3n	Corn_Nitrogen_H_to_L	Zea Mays	108532	Roots	Tissue	Tissue
				Low Nitrogen	Treatment	Compound
				16	Timepoint	Hours
3n	Corn_Nitrogen_H_to_L	Zea Mays	108548	Root	Tissue	Tissue
				Low Nitrogen	Treatment	Compound
				4	Timepoint	Hours
3m	Corn_Nitrogen_L_to_H	Zea Mays	108549	Aerial	Tissue	Tissue
				0.166	Timepoint	Hours
				Nitrogen	Treatment	Compound
3m	Corn_Nitrogen_L_to_H	Zea Mays	108550	Aerial	Tissue	Tissue
				Nitrogen	Treatment	Compound
				1.5	Timepoint	Hours
3m	Corn_Nitrogen_L_to_H	Zea Mays	108551	Aerial	Tissue	Tissue
				3	Timepoint	Hours
				Nitrogen	Treatment	Compound
3ff	Corn_RT1	Zea Mays	108599	Unknown	Plant Line	Hours
				Root	Tissue	Tissue
3p	Corn_Wounding	Zea Mays	108524	Aerial	Tissue	Tissue
				Wounding	Treatment	Compound
				1	Timepoint	Hours
3p	Corn_Wounding	Zea Mays	108525	Aerial	Tissue	Tissue
				6	Timepoint	Hours
				Wounding	Treatment	Compound
3g	Drought_Flowers	Arabidopsis	108473	Flowers	Tissue	Tissue
				7 d	Timepoint	Hours
				Drought	Treatment	Compound

3g	Drought_Flowers	Arabidopsis	108474	Flowers	Tissue	Tissue
				Drought	Treatment	Compound
				8 d (1d-post_re-watering)	Timepoint	Hours
3k	GA Treated	Arabidopsis	108484	1	Timepoint	Hours
				1	Timepoint	Hours
3k	GA Treated	Arabidopsis	108485	6	Timepoint	Hours
				6	Timepoint	Hours
3k	GA Treated	Arabidopsis	108486	12	Timepoint	Hours
				12	Timepoint	Hours
3e	Germinating Seeds	Arabidopsis	108461	Day 1	Timepoint	Hours
3e	Germinating Seeds	Arabidopsis	108462	Day 2	Timepoint	Hours
3e	Germinating Seeds	Arabidopsis	108463	Day 3	Timepoint	Hours
3e	Germinating Seeds	Arabidopsis	108464	Day 4	Timepoint	Hours
3bb	Herbicide V3.1	Arabidopsis	108465	Round up	Treatment	Compound
				12	Timepoint	Hours
3bb	Herbicide V3.1	Arabidopsis	108466	Trimec	Treatment	Compound
				12	Timepoint	Hours
3bb	Herbicide V3.1	Arabidopsis	108467	Finale	Treatment	Compound
				12	Timepoint	Hours
3bb	Herbicide V3.1	Arabidopsis	108468	Glean	Treatment	Compound
				12	Timepoint	Hours
3bb	Herbicide_v2	Arabidopsis	107871	Finale	Treatment	Compound
				4	Timepoint	Hours
3bb	Herbicide_v2	Arabidopsis	107876	Finale	Treatment	Compound
				12	Timepoint	Hours
3bb	Herbicide_v2	Arabidopsis	107881	Glean	Treatment	Compound
				4	Timepoint	Hours
3bb	Herbicide_v2	Arabidopsis	107886	Trimec	Treatment	Compound
				4	Timepoint	Hours
3bb	Herbicide_v2	Arabidopsis	107891	Trimec	Treatment	Compound
				12	Timepoint	Hours
3bb	Herbicide_v2	Arabidopsis	107896	Round-up	Treatment	Compound
				4	Timepoint	Hours
3d	Trichome Inflorescences expt	Arabidopsis	108452	Hairy Influorescence #1	Tissue	Tissue
3o	SA treatment_1	Arabidopsis	108471	Columbia	Species	Hours

	hour					
				1	Timepoint	Hours
				SA	Treatment	Compound
3o	SA treatment_1 hour	Arabidopsis	108472	CS3726	Species	Hours
				1	Timepoint	Hours
				SA	Treatment	Compound
3o	SA treatment_4 hour	Arabidopsis	108469	columbia	Species	Hours
				4	Timepoint	Hours
				SA	Treatment	Compound
3o	SA treatment_4 hour	Arabidopsis	108470	CS3726	Species	Hours
				SA	Treatment	Compound
				4	Timepoint	Hours
3o	SA treatment_AJ	Arabidopsis	107953	50	Probe Amount	% of Standard Amount
				SA	Treatment	Compound
				24	Timepoint	Hours
				Clontech	Probe Type	Probe method
3o	SA treatment_AJ	Arabidopsis	107960	50	Probe Amount	% of Standard Amount
				SA	Treatment	Compound
				24	Timepoint	Hours
				Operon	Probe Type	Probe method
3o	SA_treatment 24 hour	Arabidopsis	108443	SA	Treatment	Compound
				24	Timepoint	Hours
3o	SA_treatment 6 hour	Arabidopsis	108440	SA treatment 6 hour	Treatment	Compound
				CS3726	species	Hours
3o	SA_treatment 6 hour	Arabidopsis	108441	SA treatment 6 hour	Treatment	Compound
				Columbia	species	Hours
31	Nitrogen High transition to Low	Arabidopsis	108454	10 min	Timepoint	Hours
31	Nitrogen High transition to Low	Arabidopsis	108455	1 hr	Timepoint	Hours
3j	BR_Shoot	Arabidopsis	108478	dwf4-1	Plant Line	Hours

	Apices Expt					
3j	BR_Shoot Apices Expt	Arabidopsis	108479	AOD4-4	Plant Line	Hours
3j	BR_Shoot Apices Expt	Arabidopsis	108480	Ws-2	Plant Line	Hours
				BL	Treatment	Compound
3j	BR_Shoot Apices Expt	Arabidopsis	108481	Ws-2	Plant Line	Hours
				BRZ	Treatment	Compound
3jj	Tissue Specific Expression	Arabidopsis	108429	green flower	Tissue	Tissue
				operon	Probe Type	Probe method
				50	Probe Amount	% of Standard Amount
3jj	Tissue Specific Expression	Arabidopsis	108430	white flower	Tissue	Tissue
				50	Probe Amount	% of Standard Amount
				operon	Probe Type	Probe method
3jj	Tissue Specific Expression	Arabidopsis	108431	flowers (bud)	Tissue	Tissue
				operon	Probe Type	Probe method
				50	Probe Amount	% of Standard Amount
3c	Tissue Specific Expression	Arabidopsis	108436	5-10mm siliques	Tissue	Tissue
				33	Probe Amount	% of Standard Amount
				operon	Probe Type	Probe method
3c	Tissue Specific Expression	Arabidopsis	108437	<5mm siliques	Tissue	Tissue
				operon	Probe Type	Probe method
				33	Probe Amount	% of Standard Amount
3c	Tissue Specific	Arabidopsis	108438	5wk siliques	Tissue	Tissue

	Expression					
				33	Probe Amount	% of Standard Amount
				operon	Probe Type	Probe method
3a	Tissue Specific Expression	Arabidopsis	108439	Roots (2wk)	Tissue	Tissue
				operon	Probe Type	Probe method
				33	Probe Amount	% of Standard Amount
3c	Tissue Specific Expression	Arabidopsis	108497	3 week Rossette leaves	Tissue	Tissue
				100	Probe Amount	% of Standard Amount
				operon	Probe Type	Probe method
3c	Tissue Specific Expression	Arabidopsis	108498	3-week stems	Tissue	Tissue
				operon	Probe Type	Probe method
				100	Probe Amount	% of Standard Amount
3dd	U.A.E. Knockout	Arabidopsis	108451	13B12	Plant Line	Hours
3q	Ws Arabidopsis Drought 2 days	Arabidopsis	108477	stems and leaves	Tissue	Tissue
				2 days	Timepoint	Hours
3q	Ws Arabidopsis Drought 4 days	Arabidopsis	108482	4 days	Timepoint	Hours
3q	Ws Arabidopsis Drought 6 days	Arabidopsis	108483	6 days	Timepoint	Hours
3cc	ap2-floral buds	Arabidopsis	108501	ap2 (Ler.)	Plant Line	Hours
				floral buds	Tissue	Tissue
3m	nitrogen-seed set	Arabidopsis	108487	0.5	Timepoint	Hours
3m	nitrogen-seed set	Arabidopsis	108488	2	Timepoint	Hours
3m	nitrogen-seed set	Arabidopsis	108489	4	Timepoint	Hours

3b	rhl mutant2	Arabidopsis	108433	mutant	Tissue	Tissue
3ee	root tips	Arabidopsis	108434	root tips	Tissue	Tissue
3f	stm mutants	Arabidopsis	108435	stem	Tissue	Tissue
	Aluminum		SMD 7304, SMD 7305			
	Axel		SMD 6654, SMD 6655			
	Cadium		SMD 7427, SMD 7428			
	Cauliflower		SMD 5329, SMD 5330			
	Chloroplast		SMD 8093, SMD 8094			
	Circadian		SMD 2344, SMD 2359, SMD 2361, SMD 2362, SMD 2363, SMD 2364, SMD 2365, SMD 2366, SMD 2367, SMD 2368, SMD 3242			
	CO2		SMD7561, SMD 7562, SMD 7261, SMD			

			7263, SMD 3710, SMD 4649, SMD 4650			
	Disease		SMD 7342, SMD 7343			
	reactive oxygen		SMD 7523			
	Iron		SMD 7114, SMD 7115, SMD 7125			
	defense		SMD 8031, SMD 8032			
	Mitchondria-Electron Transport		SMD 8061, SMD 8063			
	NAA		SMD 3743, SMD 3749, SMD 6338, SMD 6339			
	Nitrogen		SMD 3787, SMD 3789			
	Phototropism		SMD 4188, SMD 6617, SMD 6619			
	Shade		SMD 8130, SMD 7230			
	Sqn		SMD 7133, SMD 7137			
	Sulfur		SMD 8034, SMD 8035			

	Wounding		SMD 3714, SMD 3715			
	Zinc		SMD 7310, SMD 7311			

6. MA_Clusters Table

Microarray data was clustered using one of two methods: "complete linkage" or "nearest neighbor" analysis. These clustering methods are described in more detail elsewhere herein. The results of the clustering analysis are presented in the MA_clust table. The table is organized as follows:

"METHOD" refers to a method number which clustering method used.

"CL_METHOD_TYPE=TRUE" refers to complete linkage method.

"NN_METHOD_TYPE=TRUE" refers to the nearest neighbor method.

"FULL_NN_METHOD_TYPE=TRUE" refers to the nearest neighbor method, where no size limitation was placed on the cluster.

"PARAMETERS" refers to the parameters utilized for the analysis. The nature of these is also described in more detail elsewhere herein.

"ORGANISM" refers to the cDNA spotted on the chip were similar to Arabidopsis thaliana (3769) sequences or whether the oligo used for the chips were similar to Zea mays (311987)sequences.

Each cluster or group of cDNA is identified by a "Group #", following which are the individual cDNA_Ids that are a member of that Group

7. Knock-in Table

The Knock-In Table presents the results of knock-in experiments wherein plants are grown from tissues transformed with a marker gene-containing insert and phenotypes are ascertained from the transformed plants. Each section of the Table relating to information on a new transformant begins with a heading "Knock-in phenotype in gene (cDNA_id):" followed by a number which represents the Ceres internal code for a proprietary cDNA sequence. The described transformant was prepared by procedures described herein, wherein the identified

Ceres proprietary cDNA_id (corresponding to the cDNA_id in the Reference and Sequence Tables) was interrupted by the marker gene-containing insert.The following information is presented for each section.

- Parent plants used in cross – presents the id numbers of the parent plants which were crossed to produce the F1 generation plant for which a phenotype is described. The parent plant with the promoter is described by a plant line descriptor.
- Clone ID – presents the clone number of the Ceres proprietary clone which was the source of the cDNA_id.
- Phenotype ID – represents an internal identification code.
- Unique FI plant ID – represents the internal code for the F1 plant for which a phenotype is described.
- Assay – presents the type of growth analyzed (e.g. soil gross morphology), followed by the assay name which corresponds to the type/location of the tissue that was obsereved., the name of the assay conducted for which the result provided the identified phenotype.
- Phenotype – describes the phenotype noted for the F1 generation transformant.
- Notes – may provide additional information on the described phenotype for the transformant.

Each knock-in representing a transformant with an interruption in the identified cDNA_id may be correlated with more than one identified phenotype.

8. Knock-out Table

The Knock-Out Table presents the results of knock-out experiments wherein plants are grown from tissues transformed with a marker gene-containing insert wherein phenotypes are ascertained from the transformed plants. Each section of the Table relating to information on a new transformant begins with a heading "tail id:" representing an internal code. The following information is presented for each section.

br - provides another internal code for the experiment.

Phenotype_ id - provides an identification number for the particular phenotype identified for the transformant.

assay - identifies the assay procedure utilized in the experiment to identify a phenotype for the transformant.

phenotype - represents an internatl identification code.

ratio - represents a segregation ratio.

notes - lists any notes relevant to the identified phenotype.

Knock-out in-genes - Identifies the genes in which the tag has inserted

6) the less than 501 upstream of the transcriptional start site;
7) less than 701 upstream of the translational initiation codon;
8) between the translational initiation and termination codons of the gene,
9) less than 301 downstream of the translational stop codon; or
10) less than 151 downstream of a transcriptional termination site or a gene.

In this table the gene is identified by its cDNA ID number, the Ceres SEQ ID that is indicated in the (Ac) portion of the Reference tables. For each cDNA_id, the following information is provided:

- the cDNA_id number.
- in parenthesis, the cluster number of which the identified cDNA is a member.
- the "gDNA_Insert pos" representing the position of the insert in the corresponding gDNA sequence
- the gi nimber refers to the TIGR chromosome sequences for Arabidopsis.

Knock-out out of-genes: Identifies the Ceres cDNA proprietary sequences (noted by cDNA_id which are the same as those identified in the Reference and Sequence Tables) which are closest in position to the insert, both upstream and downstream from the insert. For each cDNA_id, the following information is provided:

- In the first parentheses, R indicates that the gene is to the right of the tag, L indicates that the gene is to right of the tag as the sequences is read left to right
- the cDNA_id number

- in next parentheses, the cluster number of which the identified cDNA is a member.
- the distance (in number of nucleotides) of the insert is upstream of the start of the gene annotation as described in the Reference Tables or downstream at the end the gene annotation.
- the "gDNA_Insert pos" representing the position of the insert in the corresponding gDNA sequence
- the gi nimber refers to the TIGR chromosome sequences for Arabidopsis.

9. Protein Domain Table

The Protein Domain table provides details concerning the protein domains noted in the Reference Table. The majority of the protein domain descriptions given in the Protein Domain Table are obtained from Prosite, (http//www.expasy.ch/prosite/), and Pfam, (http//pfam.wustl.edu/browse.shtml). Each description in The Table begins with the pfam and Prosite identifying numbers, the full name of the domain, and a detailed description, including biological and in vivo implications/functions for the domain, references which further describe such implications/functions, and references that describe tests/assays to measure the implications/functions.

10. Single Gene Functions & Utilities Table

The Single Gene Functions & Utilities Table describes particular utilities/functions of interest for individual genes. The Table identifies the cDNA_ID of interest, correlates to that cDNA the relevant phenotype, protein domain and microarray/differential expression data. The final column of the Table identifies the utilities/functions of particular interest for the identified cDNA.

11. Cluster Functions & Utilities Table

The Cluster Functions & Utilities Table describes particular utilities/functions of interest for identified clusters of genes. The Table provides the following information:

Record # - an internal identifier.

Goup – identifies the group of clusters of interest, wherein each group is identified with the same utilities/funcions as set forth in the right-hand most column.

CDNA – identifies the cDNA of interest with the noted utility/function.

CDNA_Cluster - identifies the cDNA Cluster ID of interest.

Gi No – refers to the public genomic sequence that matches to the cDNA

NR Hit – refers to the most relevant protein domain for the cDNA of interest.

Pfam and Pfam Desc – provide the protein domain name.

Notes/Annotations – provides some notes relevant to the data/information analysis.

Utilities/Functions – this rightmost column identifies utilities/functions of particular interest for the group of cDNAs and clusters.

12. cDNA_Clusters Table

The cDNA_Clusters Table correlates the Ceres cDNA_ID nos. (in numerical order) with the relevant cDNA cluster which contains each cDNA_ID.

13. Stanford_old_new_cDNA_map Table

During the course of the experiments reported herein, some of the cDNA sequences were assigned new Ceres internal cDNA_id numbers. The cDNA_map Table provides a list of the original "old" cDNA_ids and correlates those id numbers with any new cDNA_id which may have been assigned. Thus, any "old" and "new" cDNA ids which are on the same line in the Table are, in fact, the same sequence.

14. gb_Only_Peptides Table

In the Protein Group table, a number of proteins encoded by Genbank predictions are included. These proteins were referenced with a peptide ID number. The peptide ID number is linked to the amino acid sequence of the Genbank prediction in this table.

15. Stanford_Old_New_cDNA Table

During the course of the experiments reported herein, some of the cDNA sequences utilized in the Stanford Microarray differential expression analysis experiments were assigned

new Ceres internal cDNA_id numbers. The Stanford_old_new_cDNA Table provides a list of the original "old" cDNA_ids and correlates those id numbers with any new cDNA_id which may have been assigned. Thus, any "old" and "new" cDNA ids which are on the same line in the Table are, in fact, the same sequence.

16. Enhanced_Amino Table

This table list the peptide IDs of polypeptides with enhanced amino acid content. The table list the peptide ID following with the single letter code of the amino acid that is enhanced. The table also includes a frequency that the amino acid occurred. The frequency was calculated by dividing the total number of the desired amino acid indicated in the column by the number of residues in the peptide. For example, if amino acid A, occurred 50 times in a polypeptide that is 100 amino acid long, the frequency would be 50 divided by 100 or 0.5.

17. Stanford_old_new_cDNA_map Table**Keep only for patent #3

During the course of the experiments reported herein, some of the cDNA sequences were assigned new Ceres internal cDNA_id numbers. The docket_80090_101_cDNA_map provides a list of the original "old" cDNA_ids in the Reference and Sequence tables and correlates those id numbers with any new cDNA_id which may have been assigned and utilized in the remaining tables. Thus, any "old" and "new" cDNA ids which are on the same line in the Table are, in fact, the same sequence.

17. The AFLP_int Table

The table includes the following headings:

A. Band_id
B. Mobility
C. S1-S11

A. Band_id

This text is the band identified in the experiment.

C. B. Mobility

D. The mobility of the nucleic acid bands in relation to markers in the gel.

C. S1-S11

These are the various tissue samples from Arabidopsis thaliana (wassilewskija) listed below:

S1	Dark adapted seedlings
S2	Roots/etiolated seedlings
S3	Mature leafs, soil grown
S4	Immature buds, inflorescence meristem
S5	Flowers opened
S6	Siliques, all stages
S7	Senescing leaves (just beginning to yellow)
S8	A pooled sample of:
	Callus Inducing medium
	Callus shoot induction
	Callus root induction
S9	Wounding
	Methyl-jasmonate-treated
S10	Oxidative stress
	Drought stress
	Oxygen Stress-flooding
S11	Heat treated light grown seedling
	Cold treated light grown seedlings

The numbers below each sample correspond to the levels of expression of a particular gene, polynucleotide, or nucleic acid. The stronger the intensity of the band on a AFLP gel, the higher the levels of expression.

19. The AFLP_diff Tabble

This table shows which band_id corresponds to which specific cDNA Ceres SEQ ID NOs. Differential expression of the band is shown by a "+" or "-".

HOW TO MAKE DIFFERENT EMBODIMENTS OF THE INVENTION

The invention relates to (I) polynucleotides and methods of use thereof, such as

IA. Probes, Primers and Substrates;

IB. Methods of Detection and Isolation;

B.1. Hybridization;

B.2. Methods of Mapping;

B.3. Southern Blotting;

B.4. Isolating cDNA from Related Organisms;

B.5. Isolating and/or Identifying Orthologous Genes

IC. Methods of Inhibiting Gene Expression

C.1. Antisense

C.2. Ribozyme Constructs;

C.3. Chimeraplasts;

C.4 Co-Suppression;

C.5. Transcriptional Silencing

C.6. Other Methods to Inhibit Gene Expression

ID. Methods of Functional Analysis;

IE. Promoter Sequences and Their Use;

IF. UTRs and/or Intron Sequences and Their Use; and

IG. Coding Sequences and Their Use.

The invention also relates to (II) polypeptides and proteins and methods of use thereof, such as

IIA. Native Polypeptides and Proteins

A.1 Antibodies

A.2 In Vitro Applications

IIB. Polypeptide Variants, Fragments and Fusions

B.1 Variants

B.2 Fragments

B.3 Fusions

The invention also includes (III) methods of modulating polypeptide production, such as

IIIA. Suppression

A.1 Antisense

A.2 Ribozymes

A.3 Co-suppression

A.4 Insertion of Sequences into the Gene to be Modulated

A.5 Promoter Modulation

A.6 Expression of Genes containing Dominant-Negative Mutations

IIIB. Enhanced Expression

B.1 Insertion of an Exogenous Gene

B.2 Promoter Modulation

The invention further concerns (IV) gene constructs and vector construction, such as

IVA. Coding Sequences

IVB. Promoters

IVC. Signal Peptides

The invention still further relates to

V. Transformation Techniques

I. Polynucleotides

Exemplified SDFs of the invention represent fragments of the genome of corn, wheat, rice, soybean or *Arabidopsis* and/or represent mRNA expressed from that genome. The isolated nucleic acid of the invention also encompasses corresponding fragments of the genome and/or cDNA complement of other organisms as described in detail below.

Polynucleotides of the invention can be isolated from polynucleotide libraries using primers comprising sequence similar to those described by or Reference, Sequence or polynucleotides that encode sequences Protein Group, and Protein Group Matrix tables or complements thereof. See, for example, the methods described in Sambrook *et al., supra.*

Alternatively, the polynucleotides of the invention can be produced by chemical synthesis. Such synthesis methods are described below.

It is contemplated that the nucleotide sequences presented herein may contain some small percentage of errors. These errors may arise in the normal course of determination of nucleotide sequences. Sequence errors can be corrected by obtaining seeds deposited under the accession numbers cited above, propagating them, isolating genomic DNA or appropriate mRNA from the resulting plants or seeds thereof, amplifying the relevant fragment of the genomic DNA or mRNA using primers having a sequence that flanks the erroneous sequence, and sequencing the amplification product.

I.A. Probes, Primers and Substrates

SDFs of the invention can be applied to substrates for use in array applications such as, but not limited to, assays of global gene expression, for example under varying conditions of development, growth conditions. The arrays can also be used in diagnostic or forensic methods (WO95/35505, US 5,445,943 and US 5,410,270).

Probes and primers of the instant invention will hybridize to a polynucleotide comprising a sequence in or encoded by those in the Reference, Sequence, Protein Group, and Protein Group Matrix tables or fragments or complement thereof. Though many different nucleotide sequences can encode an amino acid sequence, the sequences of the reference and Sequence table or sequences that encode polypeptides or fragments thereof described in Protein Group and Protein Group Matrix tables are generally preferred for encoding polypeptides of the invention. However, the sequence of the probes and/or primers of the instant invention need not be identical to those in the Reference and Sequence tables or the complements thereof. For example, some variation in probe or primer sequence and/or length can allow additional family members to be detected, as well as orthologous genes and more taxonomically distant related sequences. Similarly, probes and/or primers of the invention can include additional nucleotides that serve as a label for detecting the formed duplex or for subsequent cloning purposes.

Probe length will vary depending on the application. For use as primers, probes are 12-40 nucleotides, preferably 18-30 nucleotides long. For use in mapping, probes are preferably 50 to 500 nucleotides, preferably 100-250 nucleotides long. For Southern hybridizations, probes as long as several kilobases can be used as explained below.

The probes and/or primers can be produced by synthetic procedures such as the triester method of Matteucci et al. *J. Am. Chem. Soc.* 103:3185(1981); or according to Urdea et al. *Proc.*

Natl. Acad. 80:7461 (1981) or using commercially available automated oligonucleotide synthesizers.

I.B. Methods of Detection and Isolation

The polynucleotides of the invention can be utilized in a number of methods known to those skilled in the art as probes and/or primers to isolate and detect polynucleotides, including, without limitation: Southerns, Northerns, Branched DNA hybridization assays, polymerase chain reaction, and microarray assays, and variations thereof. Specific methods given by way of examples, and discussed below include:

Hybridization

Methods of Mapping

Southern Blotting

Isolating cDNA from Related Organisms

Isolating and/or Identifying Orthologous Genes.

Also, the nucleic acid molecules of the invention can used in other methods, such as high density oligonucleotide hybridizing assays, described, for example, in U.S. Pat. Nos. 6,004,753; 5,945,306; 5,945,287; 5,945,308; 5,919,686; 5,919,661; 5,919,627; 5,874,248; 5,871,973; 5,871,971; and 5,871,930; and PCT Pub. Nos. WO 9946380; WO 9933981; WO 9933870; WO 9931252; WO 9915658; WO 9906572; WO 9858052; WO 9958672; and WO 9810858.

B.1. Hybridization

The isolated SDFs of the Reference and Sequence tables or SDFs encoding polypeptides of the Protein Group and Protein Group Matrix tables or fragments thereof of the present invention can be used as probes and/or primers for detection and/or isolation of related polynucleotide sequences through hybridization. Hybridization of one nucleic acid to another constitutes a physical property that defines the subject SDF of the invention and the identified related sequences. Also, such hybridization imposes structural limitations on the pair. A good general discussion of the factors for determining hybridization conditions is provided by Sambrook et al. ("Molecular Cloning, a Laboratory Manual, 2nd ed., c. 1989 by Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; *see esp.,* chapters 11 and 12). Additional

considerations and details of the physical chemistry of hybridization are provided by G.H. Keller and M.M. Manak "DNA Probes", 2nd Ed. pp. 1-25, c. 1993 by Stockton Press, New York, NY.

Depending on the stringency of the conditions under which these probes and/or primers are used, polynucleotides exhibiting a wide range of similarity to those in the Reference and Sequence or encoding polypeptides of the Protein Group and Protein Group Matrix tables or fragments thereof can be detected or isolated. When the practitioner wishes to examine the result of membrane hybridizations under a variety of stringencies, an efficient way to do so is to perform the hybridization under a low stringency condition, then to wash the hybridization membrane under increasingly stringent conditions.

When using SDFs to identify orthologous genes in other species, the practitioner will preferably adjust the amount of target DNA of each species so that, as nearly as is practical, the same number of genome equivalents are present for each species examined. This prevents faint signals from species having large genomes, and thus small numbers of genome equivalents per mass of DNA, from erroneously being interpreted as absence of the corresponding gene in the genome.

The probes and/or primers of the instant invention can also be used to detect or isolate nucleotides that are "identical" to the probes or primers. Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below.

Isolated polynucleotides within the scope of the invention also include allelic variants of the specific sequences presented in the Reference, Sequence, Protein Group, and Protein Group Matrix tables. The probes and/or primers of the invention can also be used to detect and/or isolate polynucleotides exhibiting at least 80% sequence identity with the sequences of the reference, Sequence or encoding polypeptides of the Protein Group and Protein Group Matrix tables or fragments thereof.

With respect to nucleotide sequences, degeneracy of the genetic code provides the possibility to substitute at least one base of the base sequence of a gene with a different base without causing the amino acid sequence of the polypeptide produced from the gene to be changed. Hence, the DNA of the present invention may also have any base sequence that has been changed from a sequence in the Reference, Sequence, Protein Group, and Protein Group

Matrix tables by substitution in accordance with degeneracy of genetic code. References describing codon usage include: Carels *et al.*, *J. Mol. Evol.* 46: 45 (1998) and Fennoy *et al.*, *Nucl. Acids Res.* 21(23): 5294 (1993).

B.2. Mapping

The isolated SDF DNA of the invention can be used to create various types of genetic and physical maps of the genome of corn, Arabidopsis, soybean, rice, wheat, or other plants. Some SDFs may be absolutely associated with particular phenotypic traits, allowing construction of gross genetic maps. While not all SDFs will immediately be associated with a phenotype, all SDFs can be used as probes for identifying polymorphisms associated with phenotypes of interest. Briefly, one method of mapping involves total DNA isolation from individuals. It is subsequently cleaved with one or more restriction enzymes, separated according to mass, transferred to a solid support, hybridized with SDF DNA and the pattern of fragments compared. Polymorphisms associated with a particular SDF are visualized as differences in the size of fragments produced between individual DNA samples after digestion with a particular restriction enzyme and hybridization with the SDF. After identification of polymorphic SDF sequences, linkage studies can be conducted. By using the individuals showing polymorphisms as parents in crossing programs, F2 progeny recombinants or recombinant inbreds, for example, are then analyzed. The order of DNA polymorphisms along the chromosomes can be determined based on the frequency with which they are inherited together versus independently. The closer two polymorphisms are together in a chromosome the higher the probability that they are inherited together. Integration of the relative positions of all the polymorphisms and associated marker SDFs can produce a genetic map of the species, where the distances between markers reflect the recombination frequencies in that chromosome segment.

The use of recombinant inbred lines for such genetic mapping is described for *Arabidopsis* by Alonso-Blanco et al. (*Methods in Molecular Biology*, vol.82, "*Arabidopsis Protocols*", pp. 137-146, J.M. Martinez-Zapater and J. Salinas, eds., c. 1998 by Humana Press, Totowa, NJ) and for corn by Burr ("Mapping Genes with Recombinant Inbreds", pp. 249-254. *In* Freeling, M. and V. Walbot (Ed.), *The Maize Handbook*, c. 1994 by Springer-Verlag New York, Inc.: New York, NY, USA; Berlin Germany; Burr et al. *Genetics* (1998) 118: 519;

Gardiner, J. et al., (1993) *Genetics* 134: 917). This procedure, however, is not limited to plants and can be used for other organisms (such as yeast) or for individual cells.

The SDFs of the present invention can also be used for simple sequence repeat (SSR) mapping. Rice SSR mapping is described by Morgante et al. (*The Plant Journal* (1993) 3: 165), Panaud et al. (*Genome* (1995) 38: 1170); Senior et al. (*Crop Science* (1996) 36: 1676), Taramino et al. (*Genome* (1996) 39: 277) and Ahn et al. (*Molecular and General Genetics* (1993) 241: 483-90). SSR mapping can be achieved using various methods. In one instance, polymorphisms are identified when sequence specific probes contained within an SDF flanking an SSR are made and used in polymerase chain reaction (PCR) assays with template DNA from two or more individuals of interest. Here, a change in the number of tandem repeats between the SSR-flanking sequences produces differently sized fragments (U.S. Patent 5,766,847). Alternatively, polymorphisms can be identified by using the PCR fragment produced from the SSR-flanking sequence specific primer reaction as a probe against Southern blots representing different individuals (U.H. Refseth et al., (1997) *Electrophoresis* 18: 1519).

Genetic and physical maps of crop species have many uses. For example, these maps can be used to devise positional cloning strategies for isolating novel genes from the mapped crop species. In addition, because the genomes of closely related species are largely syntenic (that is, they display the same ordering of genes within the genome), these maps can be used to isolate novel alleles from relatives of crop species by positional cloning strategies.

The various types of maps discussed above can be used with the SDFs of the invention to identify Quantitative Trait Loci (QTLs). Many important crop traits, such as the solids content of tomatoes, are quantitative traits and result from the combined interactions of several genes. These genes reside at different loci in the genome, oftentimes on different chromosomes, and generally exhibit multiple alleles at each locus. The SDFs of the invention can be used to identify QTLs and isolate specific alleles as described by de Vicente and Tanksley (*Genetics* 134:585 (1993)). In addition to isolating QTL alleles in present crop species, the SDFs of the invention can also be used to isolate alleles from the corresponding QTL of wild relatives. Transgenic plants having various combinations of QTL alleles can then be created and the effects of the combinations measured. Once a desired allele combination has been identified, crop improvement can be accomplished either through biotechnological means or by directed

conventional breeding programs (for review *see* Tanksley and McCouch, *Science* 277:1063 (1997)).

In another embodiment, the SDFs can be used to help create physical maps of the genome of corn, *Arabidopsis* and related species. Where SDFs have been ordered on a genetic map, as described above, they can be used as probes to discover which clones in large libraries of plant DNA fragments in YACs, BACs, etc. contain the same SDF or similar sequences, thereby facilitating the assignment of the large DNA fragments to chromosomal positions. Subsequently, the large BACs, YACs, etc. can be ordered unambiguously by more detailed studies of their sequence composition (e.g. Marra et al. (1997) Genomic Research 7:1072-1084) and by using their end or other sequences to find the identical sequences in other cloned DNA fragments. The overlapping of DNA sequences in this way allows large contigs of plant sequences to be built that, when sufficiently extended, provide a complete physical map of a chromosome. Sometimes the SDFs themselves will provide the means of joining cloned sequences into a contig.

The patent publication WO95/35505 and U.S. Patents 5,445,943 and 5,410,270 describe scanning multiple alleles of a plurality of loci using hybridization to arrays of oligonucleotides. These techniques are useful for each of the types of mapping discussed above.

Following the procedures described above and using a plurality of the SDFs of the present invention, any individual can be genotyped. These individual genotypes can be used for the identification of particular cultivars, varieties, lines, ecotypes and genetically modified plants or can serve as tools for subsequent genetic studies involving multiple phenotypic traits.

B.3 Southern Blot Hybridization

The sequences from Reference and Sequence and those encoding polypeptides of Protein Group and Protein Group Matrix tables or fragments thereof can be used as probes for various hybridization techniques. These techniques are useful for detecting target polynucleotides in a sample or for determining whether transgenic plants, seeds or host cells harbor a gene or sequence of interest and thus might be expected to exhibit a particular trait or phenotype.

In addition, the SDFs from the invention can be used to isolate additional members of gene families from the same or different species and/or orthologous genes from the same or different species. This is accomplished by hybridizing an SDF to, for example, a Southern blot containing the appropriate genomic DNA or cDNA. Given the resulting hybridization data, one

of ordinary skill in the art could distinguish and isolate the correct DNA fragments by size, restriction sites, sequence and stated hybridization conditions from a gel or from a library.

Identification and isolation of orthologous genes from closely related species and alleles within a species is particularly desirable because of their potential for crop improvement. Many important crop traits, such as the solid content of tomatoes, result from the combined interactions of the products of several genes residing at different loci in the genome. Generally, alleles at each of these loci can make quantitative differences to the trait. By identifying and isolating numerous alleles for each locus from within or different species, transgenic plants with various combinations of alleles can be created and the effects of the combinations measured. Once a more favorable allele combination has been identified, crop improvement can be accomplished either through biotechnological means or by directed conventional breeding programs (Tanksley et al. *Science* 277:1063(1997)).

The results from hybridizations of the SDFs of the invention to, for example, Southern blots containing DNA from another species can also be used to generate restriction fragment maps for the corresponding genomic regions. These maps provide additional information about the relative positions of restriction sites within fragments, further distinguishing mapped DNA from the remainder of the genome.

Physical maps can be made by digesting genomic DNA with different combinations of restriction enzymes.

Probes for Southern blotting to distinguish individual restriction fragments can range in size from 15 to 20 nucleotides to several thousand nucleotides. More preferably, the probe is 100 to 1,000 nucleotides long for identifying members of a gene family when it is found that repetitive sequences would complicate the hybridization. For identifying an entire corresponding gene in another species, the probe is more preferably the length of the gene, typically 2,000 to 10,000 nucleotides, but probes 50-1,000 nucleotides long might be used. Some genes, however, might require probes up to 1,500 nucleotides long or overlapping probes constituting the full-length sequence to span their lengths.

Also, while it is preferred that the probe be homogeneous with respect to its sequence, it is not necessary. For example, as described below, a probe representing members of a gene family having diverse sequences can be generated using PCR to amplify genomic DNA or RNA templates using primers derived from SDFs that include sequences that define the gene family.

For identifying corresponding genes in another species, the next most preferable probe is a cDNA spanning the entire coding sequence, which allows all of the mRNA-coding fragment of the gene to be identified. Probes for Southern blotting can easily be generated from SDFs by making primers having the sequence at the ends of the SDF and using corn or *Arabidopsis* genomic DNA as a template. In instances where the SDF includes sequence conserved among species, primers including the conserved sequence can be used for PCR with genomic DNA from a species of interest to obtain a probe.

Similarly, if the SDF includes a domain of interest, that fragment of the SDF can be used to make primers and, with appropriate template DNA, used to make a probe to identify genes containing the domain. Alternatively, the PCR products can be resolved, for example by gel electrophoresis, and cloned and/or sequenced. Using Southern hybridization, the variants of the domain among members of a gene family, both within and across species, can be examined.

B.4.1 Isolating DNA from Related Organisms

The SDFs of the invention can be used to isolate the corresponding DNA from other organisms. Either cDNA or genomic DNA can be isolated. For isolating genomic DNA, a lambda, cosmid, BAC or YAC, or other large insert genomic library from the plant of interest can be constructed using standard molecular biology techniques as described in detail by Sambrook et al. 1989 (Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, New York) and by Ausubel et al. 1992 (Current Protocols in Molecular Biology, Greene Publishing, New York).

To screen a phage library, for example, recombinant lambda clones are plated out on appropriate bacterial medium using an appropriate *E. coli* host strain. The resulting plaques are lifted from the plates using nylon or nitrocellulose filters. The plaque lifts are processed through denaturation, neutralization, and washing treatments following the standard protocols outlined by Ausubel et al. (1992). The plaque lifts are hybridized to either radioactively labeled or non-radioactively labeled SDF DNA at room temperature for about 16 hours, usually in the presence of 50% formamide and 5X SSC (sodium chloride and sodium citrate) buffer and blocking reagents. The plaque lifts are then washed at 42°C with 1% Sodium Dodecyl Sulfate (SDS) and at a particular concentration of SSC. The SSC concentration used is dependent upon the stringency at which hybridization occurred in the initial Southern blot analysis performed. For

example, if a fragment hybridized under medium stringency (e.g., Tm - 20°C), then this condition is maintained or preferably adjusted to a less stringent condition (e.g., Tm-30°C) to wash the plaque lifts. Positive clones show detectable hybridization e.g., by exposure to X-ray films or chromogen formation. The positive clones are then subsequently isolated for purification using the same general protocol outlined above. Once the clone is purified, restriction analysis can be conducted to narrow the region corresponding to the gene of interest. The restriction analysis and succeeding subcloning steps can be done using procedures described by, for example Sambrook et al. (1989) cited above.

The procedures outlined for the lambda library are essentially similar to those used for YAC library screening, except that the YAC clones are harbored in bacterial colonies. The YAC clones are plated out at reasonable density on nitrocellulose or nylon filters supported by appropriate bacterial medium in petri plates. Following the growth of the bacterial clones, the filters are processed through the denaturation, neutralization, and washing steps following the procedures of Ausubel et al. 1992. The same hybridization procedures for lambda library screening are followed.

To isolate cDNA, similar procedures using appropriately modified vectors are employed. For instance, the library can be constructed in a lambda vector appropriate for cloning cDNA such as λgt11. Alternatively, the cDNA library can be made in a plasmid vector. cDNA for cloning can be prepared by any of the methods known in the art, but is preferably prepared as described above. Preferably, a cDNA library will include a high proportion of full-length clones.

B. 5. Isolating and/or Identifying Orthologous Genes

Probes and primers of the invention can be used to identify and/or isolate polynucleotides related to those in the Reference, Sequence, Protein Group, and Protein Group Matrix tables. Related polynucleotides are those that are native to other plant organisms and exhibit either similar sequence or encode polypeptides with similar biological activity. One specific example is an orthologous gene. Orthologous genes have the same functional activity. As such, orthologous genes may be distinguished from homologous genes. The percentage of identity is a function of evolutionary separation and, in closely related species, the percentage of identity can be 98 to 100%. The amino acid sequence of a protein encoded by an orthologous gene can be less than 75%

identical, but tends to be at least75% or at least 80% identical, more preferably at least 90%, most preferably at least 95% identical to the amino acid sequence of the reference protein.

To find orthologous genes, the probes are hybridized to nucleic acids from a species of interest under low stringency conditions, preferably one where sequences containing as much as 40-45% mismatches will be able to hybridize. This condition is established by T_m - 40°C to Tm - 48°C (*see* below). Blots are then washed under conditions of increasing stringency. It is preferable that the wash stringency be such that sequences that are 85 to 100% identical will hybridize. More preferably, sequences 90 to 100% identical will hybridize and most preferably only sequences greater than 95% identical will hybridize. One of ordinary skill in the art will recognize that, due to degeneracy in the genetic code, amino acid sequences that are identical can be encoded by DNA sequences as little as 67% identical or less. Thus, it is preferable, for example, to make an overlapping series of shorter probes, on the order of 24 to 45 nucleotides, and individually hybridize them to the same arrayed library to avoid the problem of degeneracy introducing large numbers of mismatches.

As evolutionary divergence increases, genome sequences also tend to diverge. Thus, one of skill will recognize that searches for orthologous genes between more divergent species will require the use of lower stringency conditions compared to searches between closely related species. Also, degeneracy of the genetic code is more of a problem for searches in the genome of a species more distant evolutionarily from the species that is the source of the SDF probe sequences.

Therefore the method described in Bouckaert et al., U.S. Ser. No. 60/121,700 Atty. Dkt. No. 2750-117P, Client Dkt. No. 00010.001, filed February 25, 1999, hereby incorporated in its entirety by reference, can be applied to the SDFs of the present invention to isolate related genes from plant species which do not hybridize to the corn *Arabidopsis,* soybean, rice, wheat, and other plant sequences of the reference, Sequence, Protein Group, and Protein Group Matrix tables.

Identification of the relationship of nucleotide or amino acid sequences among plant species can be done by comparing the nucleotide or amino acid sequences of SDFs of the present application with nucleotide or amino acid sequences of other SDFs such as those present in applications listed in the table below:

The SDFs of the invention can also be used as probes to search for genes that are related to the SDF within a species. Such related genes are typically considered to be members of a gene family. In such a case, the sequence similarity will often be concentrated into one or a few fragments of the sequence. The fragments of similar sequence that define the gene family typically encode a fragment of a protein or RNA that has an enzymatic or structural function. The percentage of identity in the amino acid sequence of the domain that defines the gene family is preferably at least 70%, more preferably 80 to 95%, most preferably 85 to 99%. To search for members of a gene family within a species, a low stringency hybridization is usually performed, but this will depend upon the size, distribution and degree of sequence divergence of domains that define the gene family. SDFs encompassing regulatory regions can be used to identify coordinately expressed genes by using the regulatory region sequence of the SDF as a probe.

In the instances where the SDFs are identified as being expressed from genes that confer a particular phenotype, then the SDFs can also be used as probes to assay plants of different species for those phenotypes.

I.C. Methods to Inhibit Gene Expression

The nucleic acid molecules of the present invention can be used to inhibit gene transcription and/or translation. Example of such methods include, without limitation:

Antisense Constructs;

Ribozyme Constructs;

Chimeraplast Constructs;

Co-Suppression;

Transcriptional Silencing; and

Other Methods of Gene Expression.

C.1 Antisense

In some instances it is desirable to suppress expression of an endogenous or exogenous gene. A well-known instance is the FLAVOR-SAVOR™ tomato, in which the gene encoding ACC synthase is inactivated by an antisense approach, thus delaying softening of the fruit after ripening. See for example, U.S. Patent No. 5,859,330; U.S. Patent No. 5,723,766; Oeller, et al,

Science, 254:437-439(1991); and Hamilton et al, *Nature*, 346:284-287 (1990). Also, timing of flowering can be controlled by suppression of the *FLOWERING LOCUS C* (*FLC*); high levels of this transcript are associated with late flowering, while absence of *FLC* is associated with early flowering (S.D. Michaels et al., *Plant Cell* 11:949 (1999). Also, the transition of apical meristem from production of leaves with associated shoots to flowering is regulated by *TERMINAL FLOWER1, APETALA1* and *LEAFY.* Thus, when it is desired to induce a transition from shoot production to flowering, it is desirable to suppress *TFL1* expression (S.J. Liljegren, *Plant Cell* 11:1007 (1999)). As another instance, arrested ovule development and female sterility result from suppression of the ethylene forming enzyme but can be reversed by application of ethylene (D. De Martinis et al., *Plant Cell* 11:1061 (1999)). The ability to manipulate female fertility of plants is useful in increasing fruit production and creating hybrids.

In the case of polynucleotides used to inhibit expression of an endogenous gene, the introduced sequence need not be perfectly identical to a sequence of the target endogenous gene. The introduced polynucleotide sequence will typically be at least substantially identical to the target endogenous sequence.

Some polynucleotide SDFs in the Reference, Sequence, Protein Group, and Protein Group Matrix tables represent sequences that are expressed in corn,wheat, rice, soybean *Arabidopsis* and/or other plants. Thus the invention includes using these sequences to generate antisense constructs to inhibit translation and/or degradation of transcripts of said SDFs, typically in a plant cell.

To accomplish this, a polynucleotide segment from the desired gene that can hybridize to the mRNA expressed from the desired gene (the "antisense segment") is operably linked to a promoter such that the antisense strand of RNA will be transcribed when the construct is present in a host cell. A regulated promoter can be used in the construct to control transcription of the antisense segment so that transcription occurs only under desired circumstances.

The antisense segment to be introduced generally will be substantially identical to at least a fragment of the endogenous gene or genes to be repressed. The sequence, however, need not be perfectly identical to inhibit expression. Further, the antisense product may hybridize to the untranslated region instead of or in addition to the coding sequence of the gene. The vectors of the present invention can be designed such that the inhibitory effect applies to other proteins within a family of genes exhibiting homology or substantial homology to the target gene.

For antisense suppression, the introduced antisense segment sequence also need not be full length relative to either the primary transcription product or the fully processed mRNA. Generally, a higher percentage of sequence identity can be used to compensate for the use of a shorter sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and homology of non-coding segments may be equally effective. Normally, a sequence of between about 30 or 40 nucleotides and the full length of the transcript canbe used, though a sequence of at least about 100 nucleotides is preferred, a sequence of at least about 200 nucleotides is more preferred, and a sequence of at least about 500 nucleotides is especially preferred.

C.2. Ribozymes

It is also contemplated that gene constructs representing ribozymes and based on the SDFs in the Reference and Sequence tables or those encoding polypeptides of the Protein Group and Protein Group Matrix tables and fragment thereof are an object of the invention. Ribozymes can also be used to inhibit expression of genes by suppressing the translation of the mRNA into a polypeptide. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is not itself altered, and is thus capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs.

A number of classes of ribozymes have been identified. One class of ribozymes is derived from a number of small circular RNAs, which are capable of self-cleavage and replication in plants. The RNAs replicate either alone (viroid RNAs) or with a helper virus (satellite RNAs). Examples include RNAs from avocado sunblotch viroid and the satellite RNAs from tobacco ringspot virus, lucerne transient streak virus, velvet tobacco mottle virus, solanum nodiflorum mottle virus and subterranean clover mottle virus. The design and use of target RNA-specific ribozymes is described in Haseloff et al. *Nature,* 334:585 (1988).

Like the antisense constructs above, the ribozyme sequence fragment necessary for pairing need not be identical to the target nucleotides to be cleaved, nor identical to the sequences in the Reference and Sequence tables or those encoding polypeptide of the Protein Group and Protein Group Matrix tables or fragments thereof. Ribozymes may be constructed by

combining the ribozyme sequence and some fragment of the target gene which would allow recognition of the target gene mRNA by the resulting ribozyme molecule. Generally, the sequence in the ribozyme capable of binding to the target sequence exhibits a percentage of sequence identity with at least 80%, preferably with at least 85%, more preferably with at least 90% and most preferably with at least 95%, even more preferably, with at least 96%, 97%, 98% or 99% sequence identity to some fragment of a sequence in the Reference, Sequence, Protein Group, and Protein Group Matrix tables or the complement thereof. The ribozyme can be equally effective in inhibiting mRNA translation by cleaving either in the untranslated or coding regions. Generally, a higher percentage of sequence identity can be used to compensate for the use of a shorter sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and homology of non-coding segments may be equally effective.

C.3. Chimeraplasts

The SDFs of the invention, such as those described by Reference, Sequence, Protein Group, and Protein Group Matrix tables, can also be used to construct chimeraplasts that can be introduced into a cell to produce at least one specific nucleotide change in a sequence corresponding to the SDF of the invention. A chimeraplast is an oligonucleotide comprising DNA and/or RNA that specifically hybridizes to a target region in a manner which creates a mismatched base-pair. This mismatched base-pair signals the cell's repair enzyme machinery which acts on the mismatched region resulting in the replacement, insertion or deletion of designated nucleotide(s). The altered sequence is then expressed by the cell's normal cellular mechanisms. Chimeraplasts can be designed to repair mutant genes, modify genes, introduce site-specific mutations, and/or act to interrupt or alter normal gene function (US Pat. Nos. 6,010,907 and 6,004,804; and PCT Pub. No. WO99/58723 and WO99/07865).

C.4. Sense Suppression

The SDFs of the reference, Sequence, Protein Group, and Protein Group Matrix tables of the present invention are also useful to modulate gene expression by sense suppression. Sense suppression represents another method of gene suppression by introducing at least one exogenous copy or fragment of the endogenous sequence to be suppressed.

Introduction of expression cassettes in which a nucleic acid is configured in the sense orientation with respect to the promoter into the chromosome of a plant or by a self-replicating virus has been shown to be an effective means by which to induce degradation of mRNAs of target genes. For an example of the use of this method to modulate expression of endogenous genes *see*, Napoli et al., *The Plant Cell* 2:279 (1990), and U.S. Patents Nos. 5,034,323, 5,231,020, and 5,283,184. Inhibition of expression may require some transcription of the introduced sequence.

For sense suppression, the introduced sequence generally will be substantially identical to the endogenous sequence intended to be inactivated. The minimal percentage of sequence identity will typically be greater than about 65%, but a higher percentage of sequence identity might exert a more effective reduction in the level of normal gene products. Sequence identity of more than about 80% is preferred, though about 95% to absolute identity would be most preferred. As with antisense regulation, the effect would likely apply to any other proteins within a similar family of genes exhibiting homology or substantial homology to the suppressing sequence.

C.5. Transcriptional Silencing

The nucleic acid sequences of the invention, including the SDFs of the reference, Sequence, Protein Group, and Protein Group Matrix tables, and fragments thereof, contain sequences that can be inserted into the genome of an organism resulting in transcriptional silencing. Such regulatory sequences need not be operatively linked to coding sequences to modulate transcription of a gene. Specifically, a promoter sequence without any other element of a gene can be introduced into a genome to transcriptionally silence an endogenous gene (see, for example, Vaucheret, H et al. (1998) The Plant Journal 16: 651-659). As another example, triple helices can be formed using oligonucleotides based on sequences from Reference, Sequence, Protein Group, and Protein Group Matrix tables, fragments thereof, and substantially similar sequence thereto. The oligonucleotide can be delivered to the host cell and can bind to the promoter in the genome to form a triple helix and prevent transcription. An oligonucleotide of interest is one that can bind to the promoter and block binding of a transcription factor to the promoter. In such a case, the oligonucleotide can be complementary to the sequences of the promoter that interact with transcription binding factors.

C.6. Other Methods to Inhibit Gene Expression

Yet another means of suppressing gene expression is to insert a polynucleotide into the gene of interest to disrupt transcription or translation of the gene.

Low frequency homologous recombination can be used to target a polynucleotide insert to a gene by flanking the polynucleotide insert with sequences that are substantially similar to the gene to be disrupted. Sequences from Reference, Sequence, Protein Group, and Protein Group Matrix tables, fragments thereof, and substantially similar sequence thereto can be used for homologous recombination.

In addition, random insertion of polynucleotides into a host cell genome can also be used to disrupt the gene of interest. Azpiroz-Leehan et al., *Trends in Genetics* 13:152 (1997). In this method, screening for clones from a library containing random insertions is preferred to identifying those that have polynucleotides inserted into the gene of interest. Such screening can be performed using probes and/or primers described above based on sequences from Reference, Sequence, Protein Group, and Protein Group Matrix tables, fragments thereof, and substantially similar sequence thereto. The screening can also be performed by selecting clones or R_1 plants having a desired phenotype.

I.D. Methods of Functional Analysis

The constructs described in the methods under I.C. above can be used to determine the function of the polypeptide encoded by the gene that is targeted by the constructs.

Down-regulating the transcription and translation of the targeted gene in the host cell or organisms, such as a plant, may produce phenotypic changes as compared to a wild-type cell or organism. In addition, *in vitro* assays can be used to determine if any biological activity, such as calcium flux, DNA transcription, nucleotide incorporation, etc., are being modulated by the down-regulation of the targeted gene.

Coordinated regulation of sets of genes, e.g., those contributing to a desired polygenic trait, is sometimes necessary to obtain a desired phenotype. SDFs of the invention representing transcription activation and DNA binding domains can be assembled into hybrid transcriptional activators. These hybrid transcriptional activators can be used with their corresponding DNA elements (i.e., those bound by the DNA-binding SDFs) to effect coordinated expression of desired genes (J.J. Schwarz et al., *Mol. Cell. Biol.* 12:266 (1992), A. Martinez et al., *Mol. Gen. Genet.* 261:546 (1999)).

The SDFs of the invention can also be used in the two-hybrid genetic systems to identify networks of protein-protein interactions (L. McAlister-Henn et al., *Methods* 19:330 (1999), J.C. Hu et al., *Methods* 20:80 (2000), M. Golovkin et al., *J. Biol. Chem.* 274:36428 (1999), K. Ichimura et al., *Biochem. Biophys. Res. Comm.* 253:532 (1998)). The SDFs of the invention can also be used in various expression display methods to identify important protein-DNA interactions (e.g. B. Luo et al., *J. Mol. Biol.* 266:479 (1997)).

I.E. Promoters

The SDFs of the invention are also useful as structural or regulatory sequences in a construct for modulating the expression of the corresponding gene in a plant or other organism, *e.g.* a symbiotic bacterium. For example, promoter sequences associated to SDFs of the reference, Sequence, Protein Group, and Protein Group Matrix tables of the present invention can be useful in directing expression of coding sequences either as constitutive promoters or to direct expression in particular cell types, tissues, or organs or in response to environmental stimuli.

With respect to the SDFs of the present invention a promoter is likely to be a relatively small portion of a genomic DNA (gDNA) sequence located in the first 2000 nucleotides upstream from an initial exon identified in a gDNA sequence or initial "ATG" or methionine codon or translational start site in a corresponding cDNA sequence. Such promoters are more likely to be found in the first 1000 nucleotides upstream of an initial ATG or methionine codon or translational start site of a cDNA sequence corresponding to a gDNA sequence. In particular, the promoter is usually located upstream of the transcription start site. The fragments of a particular gDNA sequence that function as elements of a promoter in a plant cell will preferably be found to hybridize to gDNA sequences presented and described in the Reference table at medium or high stringency, relevant to the length of the probe and its base composition.

Promoters are generally modular in nature. Promoters can consist of a basal promoter that functions as a site for assembly of a transcription complex comprising an RNA polymerase, for example RNA polymerase II. A typical transcription complex will include additional factors such as $TF_{II}B$, $TF_{II}D$, and $TF_{II}E$. Of these, $TF_{II}D$ appears to be the only one to bind DNA directly. The promoter might also contain one or more enhancers and/or suppressors that function as binding sites for additional transcription factors that have the function of modulating the level of transcription

with respect to tissue specificity and of transcriptional responses to particular environmental or nutritional factors, and the like.

Short DNA sequences representing binding sites for proteins can be separated from each other by intervening sequences of varying length. For example, within a particular functional module, protein binding sites may be constituted by regions of 5 to 60, preferably 10 to 30, more preferably 10 to 20 nucleotides. Within such binding sites, there are typically 2 to 6 nucleotides that specifically contact amino acids of the nucleic acid binding protein. The protein binding sites are usually separated from each other by 10 to several hundred nucleotides, typically by 15 to 150 nucleotides, often by 20 to 50 nucleotides. DNA binding sites in promoter elements often display dyad symmetry in their sequence. Often elements binding several different proteins, and/or a plurality of sites that bind the same protein, will be combined in a region of 50 to 1,000 basepairs.

Elements that have transcription regulatory function can be isolated from their corresponding endogenous gene, or the desired sequence can be synthesized, and recombined in constructs to direct expression of a coding region of a gene in a desired tissue-specific, temporal-specific or other desired manner of inducibility or suppression. When hybridizations are performed to identify or isolate elements of a promoter by hybridization to the long sequences presented in the Reference tables, conditions are adjusted to account for the above-described nature of promoters. For example short probes, constituting the element sought, are preferably used under low temperature and/or high salt conditions. When long probes, which might include several promoter elements are used, low to medium stringency conditions are preferred when hybridizing to promoters across species.

If a nucleotide sequence of an SDF, or part of the SDF, functions as a promoter or fragment of a promoter, then nucleotide substitutions, insertions or deletions that do not substantially affect the binding of relevant DNA binding proteins would be considered equivalent to the exemplified nucleotide sequence. It is envisioned that there are instances where it is desirable to decrease the binding of relevant DNA binding proteins to silence or down-regulate a promoter, or conversely to increase the binding of relevant DNA binding proteins to enhance or up-regulate a promoter and vice versa. In such instances, polynucleotides representing changes to the nucleotide sequence of the DNA-protein contact region by insertion of additional nucleotides, changes to identity of relevant nucleotides, including use of chemically-modified bases, or deletion of one or more nucleotides are considered encompassed

by the present invention. In addition, fragments of the promoter sequences described by Reference tables and variants thereof can be fused with other promoters or fragments to facilitate transcription and/or transcription in specific type of cells or under specific conditions.

Promoter function can be assayed by methods known in the art, preferably by measuring activity of a reporter gene operatively linked to the sequence being tested for promoter function. Examples of reporter genes include those encoding luciferase, green fluorescent protein, GUS, neo, cat and bar.

I.F. UTRs and Junctions

Polynucleotides comprising untranslated (UTR) sequences and intron/exon junctions are also within the scope of the invention. UTR sequences include introns and 5' or 3' untranslated regions (5' UTRs or 3' UTRs). Fragments of the sequences shown in the Reference and Sequence tables can comprise UTRs and intron/exon junctions.

These fragments of SDFs, especially UTRs, can have regulatory functions related to, for example, translation rate and mRNA stability. Thus, these fragments of SDFs can be isolated for use as elements of gene constructs for regulated production of polynucleotides encoding desired polypeptides.

Introns of genomic DNA segments might also have regulatory functions. Sometimes regulatory elements, especially transcription enhancer or suppressor elements, are found within introns. Also, elements related to stability of heteronuclear RNA and efficiency of splicing and of transport to the cytoplasm for translation can be found in intron elements. Thus, these segments can also find use as elements of expression vectors intended for use to transform plants.

Just as with promoters UTR sequences and intron/exon junctions can vary from those shown in the Reference and Sequence tables. Such changes from those sequences preferably will not affect the regulatory activity of the UTRs or intron/exon junction sequences on expression, transcription, or translation unless selected to do so. However, in some instances, down- or up-regulation of such activity may be desired to modulate traits or phenotypic or *in vitro* activity.

I.G. Coding Sequences

Isolated polynucleotides of the invention can include coding sequences that encode polypeptides comprising an amino acid sequence encoded by sequences described in the Reference and Sequence tables or an amino acid sequence presented in the Reference, Sequence, Protein Group, and Protein Group Matrix tables.

A nucleotide sequence encodes a polypeptide if a cell (or a cell free *in vitro* system) expressing that nucleotide sequence produces a polypeptide having the recited amino acid sequence when the nucleotide sequence is transcribed and the primary transcript is subsequently processed and translated by a host cell (or a cell free *in vitro* system) harboring the nucleic acid. Thus, an isolated nucleic acid that encodes a particular amino acid sequence can be a genomic sequence comprising exons and introns or a cDNA sequence that represents the product of splicing thereof. An isolated nucleic acid encoding an amino acid sequence also encompasses heteronuclear RNA, which contains sequences that are spliced out during expression, and mRNA, which lacks those sequences.

Coding sequences can be constructed using chemical synthesis techniques or by isolating coding sequences or by modifying such synthesized or isolated coding sequences as described above.

In addition to coding sequences encoding the polypeptide sequences of the reference, Sequence, Protein Group, and Protein Group Matrix tables, which are native to corn, *Arabidopsis*, soybean, rice, wheat, and other plants, the isolated polynucleotides can be polynucleotides that encode variants, fragments, and fusions of those native proteins. Such polypeptides are described below in part II.

In variant polynucleotides generally, the number of substitutions, deletions or insertions is preferably less than 20%, more preferably less than 15%; even more preferably less than 10%, 5%, 3% or 1% of the number of nucleotides comprising a particularly exemplified sequence. It is generally expected that non-degenerate nucleotide sequence changes that result in 1 to 10, more preferably 1 to 5 and most preferably 1 to 3 amino acid insertions, deletions or substitutions will not greatly affect the function of an encoded polypeptide. The most preferred embodiments are those wherein 1 to 20, preferably 1 to 10, most preferably 1 to 5 nucleotides are added to, or deleted from and/or substituted in the sequences specifically disclosed in the Reference and Sequence tables or polynucleotides that encode polypeptides of the Protein Group, and Protein Group Matrix tables or fragments thereof.

Insertions or deletions in polynucleotides intended to be used for encoding a polypeptide preferably preserve the reading frame. This consideration is not so important in instances when the polynucleotide is intended to be used as a hybridization probe.

II. Polypeptides and Proteins

IIA. Native polypeptides and proteins

Polypeptides within the scope of the invention include both native proteins as well as variants, fragments, and fusions thereof. Polypeptides of the invention are those encoded by any of the six reading frames of sequences shown in the Reference and Sequence tables, preferably encoded by the three frames reading in the 5' to 3' direction of the sequences as shown.

Native polypeptides include the proteins encoded by the sequences shown in the Reference and Sequence tables. Such native polypeptides include those encoded by allelic variants.

Polypeptide and protein variants will exhibit at least 75% sequence identity to those native polypeptides of the Reference and Sequence tables. More preferably, the polypeptide variants will exhibit at least 85% sequence identity; even more preferably, at least 90% sequence identity; more preferably at least 95%, 96%, 97%, 98%, or 99% sequence identity. Fragments of polypeptide or fragments of polypeptides will exhibit similar percentages of sequence identity to the relevant fragments of the native polypeptide. Fusions will exhibit a similar percentage of sequence identity in that fragment of the fusion represented by the variant of the native peptide.

Polypeptide and protein variants of the invention will exhibit at least 75% sequence identity to those motifs or consensus sequences of the Protein Group and Protein Group Matrix tables. More preferably, the polypeptide variants will exhibit at least 85% sequence identity; even more preferably, at least 90% sequence identity; more preferably at least 95%, 96%, 97%, 98%, or 99% sequence identity. Fragments of polypeptide or fragments of polypeptides will exhibit similar percentages of sequence identity to the relevant fragments of the native polypeptide that are indicated in the Protein Group table. Fusions will exhibit a similar percentage of sequence identity in that fragment of the fusion represented by the variant of the native peptide.

Furthermore, polypeptide variants will exhibit at least one of the functional properties of the native protein. Such properties include, without limitation, protein interaction, DNA interaction,

biological activity, immunological activity, receptor binding, signal transduction, transcription activity, growth factor activity, secondary structure, three-dimensional structure, etc. As to properties related to *in vitro* or *in vivo* activities, the variants preferably exhibit at least 60% of the activity of the native protein; more preferably at least 70%, even more preferably at least 80%, 85%, 90% or 95% of at least one activity of the native protein.

One type of variant of native polypeptides comprises amino acid substitutions, deletions and/or insertions. Conservative substitutions are preferred to maintain the function or activity of the polypeptide.

Within the scope of percentage of sequence identity described above, a polypeptide of the invention may have additional individual amino acids or amino acid sequences inserted into the polypeptide in the middle thereof and/or at the N-terminal and/or C-terminal ends thereof. Likewise, some of the amino acids or amino acid sequences may be deleted from the polypeptide.

A.1 Antibodies

Isolated polypeptides can be utilized to produce antibodies. Polypeptides of the invention can generally be used, for example, as antigens for raising antibodies by known techniques. The resulting antibodies are useful as reagents for determining the distribution of the antigen protein within the tissues of a plant or within a cell of a plant. The antibodies are also useful for examining the production level of proteins in various tissues, for example in a wild-type plant or following genetic manipulation of a plant, by methods such as Western blotting.

Antibodies of the present invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the polypeptides of the invention are first used to immunize a suitable animal, such as a mouse, rat, rabbit, or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies as detection reagents. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 μg/injection is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by *in vitro* immunization

using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization.

Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating the blood at 25°C for one hour, followed by incubating the blood at 4°C for 2-18 hours. The serum is recovered by centrifugation (e.g., 1,000xg for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

Monoclonal antibodies are prepared using the method of Kohler and Milstein, *Nature* 256: 495 (1975), or modification thereof. Typically, a mouse or rat is immunized as described above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen cells can be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate, or well, coated with the protein antigen. B-cells producing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (e.g., hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for the production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected Mab-secreting hybridomas are then cultured either *in vitro* (*e.g.*, in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

Other methods for sustaining antibody-producing B-cell clones, such as by EBV transformation, are known.

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly ^{32}P and ^{125}I), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TNB) to a blue pigment, quantifiable with a spectrophotometer.

A.2 <u>In Vitro Applications of Polypeptides</u>

Some polypeptides of the invention will have enzymatic activities that are useful *in vitro*. For example, the soybean trypsin inhibitor (Kunitz) family is one of the numerous families of

proteinase inhibitors. It comprises plant proteins which have inhibitory activity against serine proteinases from the trypsin and subtilisin families, thiol proteinases and aspartic proteinases. Thus, these peptides find *in vitro* use in protein purification protocols and perhaps in therapeutic settings requiring topical application of protease inhibitors.

Delta-aminolevulinic acid dehydratase (EC 4.2.1.24) (ALAD) catalyzes the second step in the biosynthesis of heme, the condensation of two molecules of 5-aminolevulinate to form porphobilinogen and is also involved in chlorophyll biosynthesis(Kaczor et al. (1994) Plant Physiol. 1-4: 1411-7; Smith (1988) Biochem. J. 249: 423-8; Schneider (1976) Z. naturforsch. [C] 31: 55-63). Thus, ALAD proteins can be used as catalysts in synthesis of heme derivatives. Enzymes of biosynthetic pathways generally can be used as catalysts for *in vitro* synthesis of the compounds representing products of the pathway.

Polypeptides encoded by SDFs of the invention can be engineered to provide purification reagents to identify and purify additional polypeptides that bind to them. This allows one to identify proteins that function as multimers or elucidate signal transduction or metabolic pathways. In the case of DNA binding proteins, the polypeptide can be used in a similar manner to identify the DNA determinants of specific binding (S. Pierrou et al., *Anal. Biochem.* 229:99 (1995), S. Chusacultanachai et al., *J. Biol. Chem.* 274:23591 (1999), Q. Lin et al., *J. Biol. Chem.* 272:27274 (1997)).

II.B. POLYPEPTIDE VARIANTS , FRAGMENTS, AND FUSIONS

Generally, variants , fragments, or fusions of the polypeptides encoded by the maximum length sequence(MLS) can exhibit at least one of the activities of the identified domains and/or related polypeptides described in Sections (C) and (D) of The Reference tables corresponding to the MLS of interest.

II.B .(1) Variants

A type of variant of the native polypeptides comprises amino acid substitutions. Conservative substitutions, described above (see II.), are preferred to maintain the function or activity of the polypeptide. Such substitutions include conservation of charge, polarity, hydrophobicity, size, etc. For example, one or more amino acid residues within the sequence can be substituted with another amino acid of similar polarity that acts as a functional equivalent, for

example providing a hydrogen bond in an enzymatic catalysis. Substitutes for an amino acid within an exemplified sequence are preferably made among the members of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

Within the scope of percentage of sequence identity described above, a polypeptide of the invention may have additional individual amino acids or amino acid sequences inserted into the polypeptide in the middle thereof and/or at the N-terminal and/or C-terminal ends thereof. Likewise, some of the amino acids or amino acid sequences may be deleted from the polypeptide. Amino acid substitutions may also be made in the sequences; conservative substitutions being preferred.

One preferred class of variants are those that comprise (1) the domain of an encoded polypeptide and/or (2) residues conserved between the encoded polypeptide and related polypeptides. For this class of variants, the encoded polypeptide sequence is changed by insertion, deletion, or substitution at positions flanking the domain and/or conserved residues.

Another class of variants includes those that comprise an encoded polypeptide sequence that is changed in the domain or conserved residues by a conservative substitution.

Yet another class of variants includes those that lack one of the *in vitro* activities, or structural features of the encoded polypeptides. One example is polypeptides or proteins produced from genes comprising dominant negative mutations. Such a variant may comprise an encoded polypeptide sequence with non-conservative changes in a particular domain or group of conserved residues.

II.A.(2) FRAGMENTS

Fragments of particular interest are those that comprise a domain identified for a polypeptide encoded by an MLS of the instant invention and variants thereof. Also, fragments that comprise at least one region of residues conserved between an MLS encoded polypeptide and its related polypeptides are of great interest. Fragments are sometimes useful as polypeptides corresponding to genes comprising dominant negative mutations are.

II.A.(3) FUSIONS

Of interest are chimeras comprising (1) a fragment of the MLS encoded polypeptide or variants thereof of interest and (2) a fragment of a polypeptide comprising the same domain. For example, an AP2 helix encoded by a MLS of the invention fused to second AP2 helix from ANT protein, which comprises two AP2 helices. The present invention also encompasses fusions of MLS encoded polypeptides, variants, or fragments thereof fused with related proteins or fragments thereof.

DEFINITION OF DOMAINS

The polypeptides of the invention may possess identifying domains as shown in The Reference tables. Specific domains within the MLS encoded polypeptides are indicated in The Reference tables. In addition, the domains within the MLS encoded polypeptide can be defined by the region that exhibits at least 70% sequence identity with the consensus sequences listed in the detailed description below of each of the domains.

The majority of the protein domain descriptions given in the protein domain table are obtained from Prosite, (http//www.expasy.ch/prosite/), and Pfam, (http//pfam.wustl.edu/browse.shtml). Examples of domain descriptions are listed in the Protein Domain table.

A. Activities of Polypeptides Comprising Signal Peptides

Polypeptides comprising signal peptides are a family of proteins that are typically targeted to (1) a particular organelle or intracellular compartment, (2) interact with a particular molecule or (3) for secretion outside of a host cell. Example of polypeptides comprising signal peptides include, without limitation, secreted proteins, soluble proteins, receptors, proteins retained in the ER, etc.

These proteins comprising signal peptides are useful to modulate ligand-receptor interactions, cell-to-cell communication, signal transduction, intracellular communication, and activities and/or chemical cascades that take part in an organism outside or within of any particular cell.

One class of such proteins are soluble proteins which are transported out of the cell. These proteins can act as ligands that bind to receptor to trigger signal transduction or to permit communication between cells.

Another class is receptor proteins which also comprise a retention domain that lodges the receptor protein in the membrane when the cell transports the receptor to the surface of the cell. Like the soluble ligands, receptors can also modulate signal transduction and communication between cells.

In addition the signal peptide itself can serve as a ligand for some receptors. An example is the interaction of the ER targeting signal peptide with the signal recognition particle (SRP). Here, the SRP binds to the signal peptide, halting translation, and the resulting SRP complex then binds to docking proteins located on the surface of the ER, prompting transfer of the protein into the ER.

A description of signal peptide residue composition is described below in Subsection IV.C.1.

III. Methods of Modulating Polypeptide Production

It is contemplated that polynucleotides of the invention can be incorporated into a host cell or in-vitro system to modulate polypeptide production. For instance, the SDFs prepared as described herein can be used to prepare expression cassettes useful in a number of techniques for suppressing or enhancing expression.

An example are polynucleotides comprising sequences to be transcribed, such as coding sequences, of the present invention can be inserted into nucleic acid constructs to modulate polypeptide production. Typically, such sequences to be transcribed are heterologous to at least one element of the nucleic acid construct to generate a chimeric gene or construct.

Another example of useful polynucleotides are nucleic acid molecules comprising regulatory sequences of the present invention. Chimeric genes or constructs can be generated

when the regulatory sequences of the invention linked to heterologous sequences in a vector construct. Within the scope of invention are such chimeric gene and/or constructs.

Also within the scope of the invention are nucleic acid molecules, whereof at least a part or fragment of these DNA molecules are presented in the Reference and Sequence tables or polynucleotide encoding polypeptides of the Protein Group or Protein Group Matrix tables of the present application, and wherein the coding sequence is under the control of its own promoter and/or its own regulatory elements. Such molecules are useful for transforming the genome of a host cell or an organism regenerated from said host cell for modulating polypeptide production.

Additionally, a vector capable of producing the oligonucleotide can be inserted into the host cell to deliver the oligonucleotide.

More detailed description of components to be included in vector constructs are described both above and below.

Whether the chimeric vectors or native nucleic acids are utilized, such polynucleotides can be incorporated into a host cell to modulate polypeptide production. Native genes and/or nucleic acid molecules can be effective when exogenous to the host cell.

Methods of modulating polypeptide expression includes, without limitation:

Suppression methods, such as

Antisense

Ribozymes

Co-suppression

Insertion of Sequences into the Gene to be Modulated

Regulatory Sequence Modulation.

as well as Methods for Enhancing Production, such as

Insertion of Exogenous Sequences; and

Regulatory Sequence Modulation.

III.A. Suppression

Expression cassettes of the invention can be used to suppress expression of endogenous genes which comprise the SDF sequence. Inhibiting expression can be useful, for instance, to

tailor the ripening characteristics of a fruit (Oeller et al., *Science* 254:437 (1991)) or to influence seed size (WO98/07842) or to provoke cell ablation (Mariani et al., Nature 357: 384-387 (1992).

As described above, a number of methods can be used to inhibit gene expression in plants, such as antisense, ribozyme, introduction of exogenous genes into a host cell, insertion of a polynucleotide sequence into the coding sequence and/or the promoter of the endogenous gene of interest, and the like.

III.A.1. Antisense

An expression cassette as described above can be transformed into host cell or plant to produce an antisense strand of RNA. For plant cells, antisense RNA inhibits gene expression by preventing the accumulation of mRNA which encodes the enzyme of interest, *see*, e.g., Sheehy et al., *Proc. Nat. Acad. Sci.* USA, 85:8805 (1988), and Hiatt et al., U.S. Patent No. 4,801,340.

III.A.2. Ribozymes

Similarly, ribozyme constructs can be transformed into a plant to cleave mRNA and down-regulate translation.

III.A.3. Co-Suppression

Another method of suppression is by introducing an exogenous copy of the gene to be suppressed. Introduction of expression cassettes in which a nucleic acid is configured in the sense orientation with respect to the promoter has been shown to prevent the accumulation of mRNA. A detailed description of this method is described above.

III.A.4. Insertion of Sequences into the Gene to be Modulated

Yet another means of suppressing gene expression is to insert a polynucleotide into the gene of interest to disrupt transcription or translation of the gene.

Homologous recombination could be used to target a polynucleotide insert to a gene using the Cre-Lox system (A.C. Vergunst et al., *Nucleic Acids Res.* 26:2729 (1998), A.C. Vergunst et al., *Plant Mol. Biol.* 38:393 (1998), H. Albert et al., *Plant J.* 7:649 (1995)).

In addition, random insertion of polynucleotides into a host cell genome can also be used to disrupt the gene of interest. Azpiroz-Leehan et al., *Trends in Genetics* 13:152 (1997). In this

method, screening for clones from a library containing random insertions is preferred for identifying those that have polynucleotides inserted into the gene of interest. Such screening can be performed using probes and/or primers described above based on sequences from the Reference and Sequence tables or polynucleotides encoding polypeptides of the Protein Group or Protein Group Matrix tables, fragments thereof, and substantially similar sequence thereto. The screening can also be performed by selecting clones or any transgenic plants having a desired phenotype.

III.A.5. Regulatory SequenceModulation

The SDFs described in the Reference and Sequence tables or polynucleotides encoding polypeptides of the Protein Group or Protein Group Matrix tables, and fragments thereof are examples of nucleotides of the invention that contain regulatory sequences that can be used to suppress or inactivate transcription and/or translation from a gene of interest as discussed in I.C.5.

III.A.6. Genes Comprising Dominant-Negative Mutations

When suppression of production of the endogenous, native protein is desired it is often helpful to express a gene comprising a dominant negative mutation. Production of protein variants produced from genes comprising dominant negative mutations is a useful tool for research Genes comprising dominant negative mutations can produce a variant polypeptide which is capable of competing with the native polypeptide, but which does not produce the native result. Consequently, over expression of genes comprising these mutations can titrate out an undesired activity of the native protein. For example, The product from a gene comprising a dominant negative mutation of a receptor can be used to constitutively activate or suppress a signal transduction cascade, allowing examination of the phenotype and thus the trait(s) controlled by that receptor and pathway. Alternatively, the protein arising from the gene comprising a dominant-negative mutation can be an inactive enzyme still capable of binding to the same substrate as the native protein and therefore competes with such native protein.

Products from genes comprising dominant-negative mutations can also act upon the native protein itself to prevent activity. For example, the native protein may be active only as a homo-multimer or as one subunit of a hetero-multimer. Incorporation of an inactive subunit into the multimer with native subunit(s) can inhibit activity.

Thus, gene function can be modulated in host cells of interest by insertion into these cells vector constructs comprising a gene comprising a dominant-negative mutation.

III.B. Enhanced Expression

Enhanced expression of a gene of interest in a host cell can be accomplished by either (1) insertion of an exogenous gene; or (2) promoter modulation.

III.B.1. Insertion of an Exogenous Gene

Insertion of an expression construct encoding an exogenous gene can boost the number of gene copies expressed in a host cell.

Such expression constructs can comprise genes that either encode the native protein that is of interest or that encode a variant that exhibits enhanced activity as compared to the native protein. Such genes encoding proteins of interest can be constructed from the sequences from the Reference and Sequence tables or polynucleotides encoding polypeptides of the Protein Group or Protein Group Matrix tables, fragments thereof, and substantially similar sequence thereto.

Such an exogenous gene can include either a constitutive promoter permitting expression in any cell in a host organism or a promoter that directs transcription only in particular cells or times during a host cell life cycle or in response to environmental stimuli.

III.B.2. Regulatory Sequence Modulation

The SDFs of the Reference and Sequence tables, and fragments thereof, contain regulatory sequences that can be used to enhance expression of a gene of interest. For example, some of these sequences contain useful enhancer elements. In some cases, duplication of enhancer elements or insertion of exogenous enhancer elements will increase expression of a desired gene from a particular promoter. As other examples, all ll promoters require binding of a regulatory protein to be activated, while some promoters may need a protein that signals a promoter binding protein to expose a polymerase binding site. In either case, over-production of such proteins can be used to enhance expression of a gene of interest by increasing the activation time of the promoter.

Such regulatory proteins are encoded by some of the sequences in the Reference and Sequence tables or polynucleotides encoding polypeptides of the Protein Group or Protein Group Matrix tables, fragments thereof, and substantially similar sequences thereto.

Coding sequences for these proteins can be constructed as described above.

IV. Gene Constructs and Vector Construction

To use isolated SDFs of the present invention or a combination of them or parts and/or mutants and/or fusions of said SDFs in the above techniques, recombinant DNA vectors which comprise said SDFs and are suitable for transformation of cells, such as plant cells, are usually prepared. The SDF construct can be made using standard recombinant DNA techniques (Sambrook et al. 1989) and can be introduced to the species of interest by *Agrobacterium*-mediated transformation or by other means of transformation (*e.g.*, particle gun bombardment) as referenced below.

The vector backbone can be any of those typical in the art such as plasmids, viruses, artificial chromosomes, BACs, YACs and PACs and vectors of the sort described by

(a) **BAC:** Shizuya et al., Proc. Natl. Acad. Sci. USA 89: 8794-8797 (1992); Hamilton et al., Proc. Natl. Acad. Sci. USA 93: 9975-9979 (1996);

(b) **YAC:** Burke et al., Science 236:806-812 (1987);.

(c) **PAC:** Sternberg N. et al., Proc Natl Acad Sci U S A. Jan;87(1):103-7 (1990);

(d) **Bacteria-Yeast Shuttle Vectors:** Bradshaw et al., Nucl Acids Res 23: 4850-4856 (1995);

(e) **Lambda Phage Vectors**: Replacement Vector, e.g., Frischauf et al., J. Mol Biol 170: 827-842 (1983); or Insertion vector, e.g., Huynh et al., In: Glover NM (ed) DNA Cloning: A practical Approach, Vol.1 Oxford: IRL Press (1985);

(f) **T-DNA gene fusion vectors :**Walden et al., Mol Cell Biol 1: 175-194 (1990); and

(g) **Plasmid vectors:** Sambrook et al., infra.

Typically, a vector will comprise the exogenous gene, which in its turn comprises an SDF of the present invention to be introduced into the genome of a host cell, and which gene may be an antisense construct, a ribozyme construct chimeraplast, or a coding sequence with any desired transcriptional and/or translational regulatory sequences, such as promoters, UTRs, and 3' end termination sequences. Vectors of the invention can also include origins of replication, scaffold attachment regions (SARs), markers, homologous sequences, introns, etc.

A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding a full length protein, will preferably be combined with transcriptional and translational initiation regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant.

For example, for over-expression, a plant promoter fragment may be employed that will direct transcription of the gene in all tissues of a regenerated plant. Alternatively, the plant promoter may direct transcription of an SDF of the invention in a specific tissue (tissue-specific promoters) or may be otherwise under more precise environmental control (inducible promoters).

If proper polypeptide productionis desired, a polyadenylation region at the 3'-end of the coding region is typically included. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA.

The vector comprising the sequences from genes or SDF or the invention may comprise a marker gene that confers a selectable phenotype on plant cells. The vector can include promoter and coding sequence, for instance. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or phosphinotricin.

IV.A. Coding Sequences

Generally, the sequence in the transformation vector and to be introduced into the genome of the host cell does not need to be absolutely identical to an SDF of the present invention. Also, it is not necessary for it to be full length, relative to either the primary transcription product or fully processed mRNA. Furthermore, the introduced sequence need not have the same intron or exon pattern as a native gene. Also, heterologous non-coding segments can be incorporated into the coding sequence without changing the desired amino acid sequence of the polypeptide to be produced.

IV.B. Promoters

As explained above, introducing an exogenous SDF from the same species or an orthologous SDF from another species are useful to modulate the expression of a native gene corresponding to that SDF of interest. Such an SDF construct can be under the control of either a constitutive promoter or a highly regulated inducible promoter (*e.g.*, a copper inducible

promoter). The promoter of interest can initially be either endogenous or heterologous to the species in question. When re-introduced into the genome of said species, such promoter becomes exogenous to said species. Over-expression of an SDF transgene can lead to co-suppression of the homologous endogeneous sequence thereby creating some alterations in the phenotypes of the transformed species as demonstrated by similar analysis of the chalcone synthase gene (Napoli et al., *Plant Cell* 2:279 (1990) and van der Krol et al., *Plant Cell* 2:291 (1990)). If an SDF is found to encode a protein with desirable characteristics, its over-production can be controlled so that its accumulation can be manipulated in an organ- or tissue-specific manner utilizing a promoter having such specificity.

Likewise, if the promoter of an SDF (or an SDF that includes a promoter) is found to be tissue-specific or developmentally regulated, such a promoter can be utilized to drive or facilitate the transcription of a specific gene of interest (*e.g.*, seed storage protein or root-specific protein). Thus, the level of accumulation of a particular protein can be manipulated or its spatial localization in an organ- or tissue- specific manner can be altered.

IV. C Signal Peptides

SDFs of the present invention containing signal peptides are indicated in the Reference, Sequence, the Protein Group and Protein Group Matrix tables. In some cases it may be desirable for the protein encoded by an introduced exogenous or orthologous SDF to be targeted (1) to a particular organelle intracellular compartment, (2) to interact with a particular molecule such as a membrane molecule or (3) for secretion outside of the cell harboring the introduced SDF. This will be accomplished using a signal peptide.

Signal peptides direct protein targeting, are involved in ligand-receptor interactions and act in cell to cell communication. Many proteins, especially soluble proteins, contain a signal peptide that targets the protein to one of several different intracellular compartments. In plants, these compartments include, but are not limited to, the endoplasmic reticulum (ER), mitochondria, plastids (such as chloroplasts), the vacuole, the Golgi apparatus, protein storage vessicles (PSV) and, in general, membranes. Some signal peptide sequences are conserved, such as the Asn-Pro-Ile-Arg amino acid motif found in the N-terminal propeptide signal that targets proteins to the vacuole (Marty (1999) *The Plant Cell* 11: 587-599). Other signal peptides do not have a consensus sequence *per se*, but are largely composed of hydrophobic amino acids, such as

those signal peptides targeting proteins to the ER (Vitale and Denecke (1999) *The Plant Cell* 11: 615-628). Still others do not appear to contain either a consensus sequence or an identified common secondary sequence, for instance the chloroplast stromal targeting signal peptides (Keegstra and Cline (1999) *The Plant Cell* 11: 557-570). Furthermore, some targeting peptides are bipartite, directing proteins first to an organelle and then to a membrane within the organelle (e.g. within the thylakoid lumen of the chloroplast; see Keegstra and Cline (1999) *The Plant Cell* 11: 557-570). In addition to the diversity in sequence and secondary structure, placement of the signal peptide is also varied. Proteins destined for the vacuole, for example, have targeting signal peptides found at the N-terminus, at the C-terminus and at a surface location in mature, folded proteins. Signal peptides also serve as ligands for some receptors.

These characteristics of signal proteins can be used to more tightly control the phenotypic expression of introduced SDFs. In particular, associating the appropriate signal sequence with a specific SDF can allow sequestering of the protein in specific organelles (plastids, as an example), secretion outside of the cell, targeting interaction with particular receptors, etc. Hence, the inclusion of signal proteins in constructs involving the SDFs of the invention increases the range of manipulation of SDF phenotypic expression. The nucleotide sequence of the signal peptide can be isolated from characterized genes using common molecular biological techniques or can be synthesized in vitro.

In addition, the native signal peptide sequences, both amino acid and nucleotide, described in the Reference, Sequence, Protein Group or Protein Group Matrix tables can be used to modulate polypeptide transport. Further variants of the native signal peptides described in the Reference, Sequence, Protein Group or Protein Group Matrix tables are contemplated. Insertions, deletions, or substitutions can be made. Such variants will retain at least one of the functions of the native signal peptide as well as exhibiting some degree of sequence identity to the native sequence.

Also, fragments of the signal peptides of the invention are useful and can be fused with other signal peptides of interest to modulate transport of a polypeptide.

V. Transformation Techniques

A wide range of techniques for inserting exogenous polynucleotides are known for a number of host cells, including, without limitation, bacterial, yeast, mammalian, insect and plant cells.

Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature. *See, e.g.* Weising et al., *Ann. Rev. Genet.* 22:421 (1988); and Christou, Euphytica, v. 85, n.1-3:13-27, (1995).

DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria (McCormac et al., *Mol. Biotechnol.* 8:199 (1997); Hamilton, *Gene* 200:107 (1997)); Salomon et al. *EMBO J.* 3:141 (1984); Herrera-Estrella et al. *EMBO J.* 2:987 (1983).

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski et al. *EMBO J.* 3:2717 (1984). Electroporation techniques are described in Fromm et al. *Proc. Natl Acad. Sci. USA* 82:5824 (1985). Ballistic transformation techniques are described in Klein et al. *Nature* 327:773 (1987). *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary or co-integrate vectors, are well described in the scientific literature. See, for example Hamilton, *CM.*, *Gene* 200:107 (1997); Müller et al. *Mol. Gen. Genet.* 207:171 (1987); Komari et al. *Plant J.* 10:165 (1996); Venkateswarlu et al. *Biotechnology* 9:1103 (1991) and Gleave, *AP.*, *Plant Mol. Biol.* 20:1203 (1992); Graves and Goldman, *Plant Mol. Biol.* 7:34 (1986) and Gould et al., *Plant Physiology* 95:426 (1991).

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant that possesses the transformed genotype and thus the desired phenotype such as seedlessness. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or

herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans et al., *Protoplasts Isolation and Culture* in "Handbook of Plant Cell Culture," pp. 124-176, MacMillan Publishing Company, New York, 1983; and Binding, *Regeneration of Plants, Plant Protoplasts,* pp. 21-73, CRC Press, Boca Raton, 1988. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee et al. *Ann. Rev. of Plant Phys.* 38:467 (1987). Regeneration of monocots (rice) is described by Hosoyama et al. (*Biosci. Biotechnol. Biochem.* 58:1500 (1994)) and by Ghosh et al. (*J. Biotechnol.* 32:1 (1994)). The nucleic acids of the invention can be used to confer desired traits on essentially any plant.

Thus, the invention has use over a broad range of plants, including species from the genera *Anacardium, Arachis, Asparagus, Atropa, Avena, Brassica, Citrus, Citrullus, Capsicum, Carthamus, Cocos, Coffea, Cucumis, Cucurbita, Daucus, Elaeis, Fragaria, Glycine, Gossypium, Helianthus, Heterocallis, Hordeum, Hyoscyamus, Lactuca, Linum, Lolium,Lupinus, Lycopersicon, Malus, Manihot, Majorana, Medicago, Nicotiana, Olea, Oryza, Panieum, Pannesetum, Persea, Phaseolus, Pistachia, Pisum, Pyrus, Prunus, Raphanus, Ricinus, Secale, Senecio, Sinapis, Solanum, Sorghum, Theobromus, Trigonella, Triticum, Vicia, Vitis, Vigna, and, Zea.*

One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

The particular sequences of SDFs identified are provided in the attached Reference and Sequence tables.

DEFINITIONS

Allelic variant: An "allelic variant" is an alternative form of the same SDF, which resides at the same chromosomal locus in the organism. Allelic variations can occur in any portion of the gene sequence, including regulatory regions. Allelic variants can arise by normal genetic variation in a population. Allelic variants can also be produced by genetic engineering methods. An allelic variant can be one that is found in a naturally occurring plant, including a cultivar or ecotype. An allelic variant may or may not give rise to a phenotypic change, and may or may not be expressed. An allele can result in a detectable change in the phenotype of the trait represented by the locus. A phenotypically silent allele can give rise to a product.

Alternatively spliced messages: Within the context of the current invention, "alternatively spliced messages" refers to mature mRNAs originating from a single gene with variations in the number and/or identity of exons, introns and/or intron-exon junctions.

Chimeric: The term "chimeric" is used to describe genes, as defined supra, or contructs wherein at least two of the elements of the gene or construct, such as the promoter and the coding sequence and/or other regulatory sequences and/or filler sequences and/or complements thereof, are heterologous to each other.

Constitutive Promoter: Promoters referred to herein as "constitutive promoters" actively promote transcription under most, but not necessarily all, environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcript initiation region and the 1' or 2' promoter derived from T-DNA of *Agrobacterium tumefaciens,* and other transcription initiation regions from various plant genes, such as the maize ubiquitin-1 promoter, known to those of skill.

Coordinately Expressed: The term "coordinately expressed," as used in the current invention, refers to genes that are expressed at the same or a similar time and/or stage and/or under the same or similar environmental conditions.

Domain: Domains are fingerprints or signatures that can be used to characterize protein families and/or parts of proteins. Such fingerprints or signatures can comprise conserved (1) primary sequence, (2) secondary structure, and/or (3) three-dimensional conformation. Generally, each domain has been associated with either a family of proteins or motifs. Typically, these families and/or motifs have been correlated with specific *in-vitro* and/or *in-vivo* activities. A domain can be any length, including the entirety of the sequence of a protein. Detailed descriptions of the domains, associated families and motifs, and correlated activities of the polypeptides of the instant invention are described below. Usually, the polypeptides with designated domain(s) can exhibit at least one activity that is exhibited by any polypeptide that comprises the same domain(s).

Endogenous: The term "endogenous," within the context of the current invention refers to any polynucleotide, polypeptide or protein sequence which is a natural part of a cell or organisms regenerated from said cell.

Exogenous: "Exogenous," as referred to within, is any polynucleotide, polypeptide or protein sequence, whether chimeric or not, that is initially or subsequently introduced into the genome of an individual host cell or the organism regenerated from said host cell by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation (of dicots - *e.g.* Salomon et al. *EMBO J.* 3:141 (1984); Herrera-Estrella et al. *EMBO J.* 2:987 (1983); of monocots, representative papers are those by Escudero et al., *Plant J.* 10:355 (1996), Ishida et al., *Nature Biotechnology* 14:745 (1996), May et al., *Bio/Technology* 13:486 (1995)), biolistic methods (Armaleo et al., *Current Genetics* 17:97 1990)), electroporation, *in planta* techniques, and the like. Such a plant containing the exogenous nucleic acid is referred to here as a T_0 for the primary transgenic plant and T_1 for the first generation. The term "exogenous" as used herein is also intended to encompass inserting a naturally found element into a non-naturally found location.

Filler sequence: As used herein, "filler sequence" refers to any nucleotide sequence that is inserted into DNA construct to evoke a particular spacing between particular components such as

a promoter and a coding region and may provide an additional attribute such as a restriction enzyme site.

Gene: The term “gene,” as used in the context of the current invention, encompasses all regulatory and coding sequence contiguously associated with a single hereditary unit with a genetic function (see SCHEMATIC 1). Genes can include non-coding sequences that modulate the genetic function that include, but are not limited to, those that specify polyadenylation, transcriptional regulation, DNA conformation, chromatin conformation, extent and position of base methylation and binding sites of proteins that control all of these. Genes comprised of “exons” (coding sequences), which may be interrupted by “introns” (non-coding sequences), encode proteins. A gene’s genetic function may require only RNA expression or protein production, or may only require binding of proteins and/or nucleic acids without associated expression. In certain cases, genes adjacent to one another may share sequence in such a way that one gene will overlap the other. A gene can be found within the genome of an organism, artificial chromosome, plasmid, vector, etc., or as a separate isolated entity.

Gene Family: “Gene family” is used in the current invention to describe a group of functionally related genes, each of which encodes a separate protein.

Heterologous sequences: “Heterologous sequences” are those that are not operatively linked or are not contiguous to each other in nature. For example, a promoter from corn is considered heterologous to an *Arabidopsis* coding region sequence. Also, a promoter from a gene encoding a growth factor from corn is considered heterologous to a sequence encoding the corn receptor for the growth factor. Regulatory element sequences, such as UTRs or 3’ end termination sequences that do not originate in nature from the same gene as the coding sequence originates from, are considered heterologous to said coding sequence. Elements operatively linked in nature and - contiguous to each other are not heterologous to each other. On the other hand, these same elements remain operatively linked but become heterologous if other filler sequence is placed between them. Thus, the promoter and coding sequences of a corn gene expressing an amino acid transporter are not heterologous to each other, but the promoter and coding sequence of a corn gene operatively linked in a novel manner are heterologous.

Homologous gene: In the current invention, "homologous gene" refers to a gene that shares sequence similarity with the gene of interest. This similarity may be in only a fragment of the sequence and often represents a functional domain such as, examples including without limitation a DNA binding domain, a domain with tyrosine kinase activity, or the like. The functional activities of homologous genes are not necessarily the same.

Inducible Promoter: An "inducible promoter" in the context of the current invention refers to a promoter which is regulated under certain conditions, such as light, chemical concentration, protein concentration, conditions in an organism, cell, or organelle, etc. A typical example of an inducible promoter, which can be utilized with the polynucleotides of the present invention, is PARSK1, the promoter from the *Arabidopsis* gene encoding a serine-threonine kinase enzyme, and which promoter is induced by dehydration, abscissic acid and sodium chloride (Wang and Goodman, *Plant J.* 8:37 (1995)) Examples of environmental conditions that may affect transcription by inducible promoters include anaerobic conditions, elevated temperature, or the presence of light.

Intergenic region: "Intergenic region," as used in the current invention, refers to nucleotide sequence occurring in the genome that separates adjacent genes.

Mutant gene: In the current invention, "mutant" refers to a heritable change in DNA sequence at a specific location. Mutants of the current invention may or may not have an associated identifiable function when the mutant gene is transcribed.

Orthologous Gene: In the current invention "orthologous gene" refers to a second gene that encodes a gene product that performs a similar function as the product of a first gene. The orthologous gene may also have a degree of sequence similarity to the first gene. The orthologous gene may encode a polypeptide that exhibits a degree of sequence similarity to a polypeptide corresponding to a first gene. The sequence similarity can be found within a functional domain or along the entire length of the coding sequence of the genes and/or their corresponding polypeptides.

Percentage of sequence identity: "Percentage of sequence identity," as used herein, is determined by comparing two optimally aligned sequences over a comparison window, where the fragment of the polynucleotide or amino acid sequence in the comparison window may comprise additions or deletions (e.g., gaps or overhangs) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman *Add. APL. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman and Wunsch *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson and Lipman *Proc. Natl. Acad. Sci. (USA)* 85: 2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, PASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection. Given that two sequences have been identified for comparison, GAP and BESTFIT are preferably employed to determine their optimal alignment. Typically, the default values of 5.00 for gap weight and 0.30 for gap weight length are used. The term "substantial sequence identity" between polynucleotide or polypeptide sequences refers to polynucleotide or polypeptide comprising a sequence that has at least 80% sequence identity, preferably at least 85%, more preferably at least 90% and most preferably at least 95%, even more preferably, at least 96%, 97%, 98% or 99% sequence identity compared to a reference sequence using the programs.

Plant Promoter: A “plant promoter” is a promoter capable of initiating transcription in plant cells and can drive or facilitate transcription of a fragment of the SDF of the instant invention or a coding sequence of the SDF of the instant invention. Such promoters need not be of plant origin. For example, promoters derived from plant viruses, such as the CaMV35S promoter or from *Agrobacterium tumefaciens* such as the T-DNA promoters, can be plant promoters. A typical example of a plant promoter of plant origin is the maize ubiquitin-1 (ubi-1)promoter known to those of skill.

Promoter: The term "promoter," as used herein, refers to a region of sequence determinants located upstream from the start of transcription of a gene and which are involved in recognition and binding of RNA polymerase and other proteins to initiate and modulate transcription. A basal promoter is the minimal sequence necessary for assembly of a transcription complex required for transcription initiation. Basal promoters frequently include a "TATA box" element usually located between 15 and 35 nucleotides upstream from the site of initiation of transcription. Basal promoters also sometimes include a "CCAAT box" element (typically a sequence CCAAT) and/or a GGGCG sequence, usually located between 40 and 200 nucleotides, preferably 60 to 120 nucleotides, upstream from the start site of transcription.

Public sequence: The term "public sequence ," as used in the context of the instant application, refers to any sequence that has been deposited in a publicly accessible database. This term encompasses both amino acid and nucleotide sequences. Such sequences are publicly accessible, for example, on the BLAST databases on the NCBI FTP web site (accessible at ncbi.nlm.gov/blast). The database at the NCBI GTP site utilizes "gi" numbers assigned by NCBI as a unique identifier for each sequence in the databases, thereby providing a non-redundant database for sequence from various databases, including GenBank, EMBL, DBBJ, (DNA Database of Japan) and PDB (Brookhaven Protein Data Bank).

Regulatory Sequence: The term "regulatory sequence," as used in the current invention, refers to any nucleotide sequence that influences transcription or translation initiation and rate, and stability and/or mobility of the transcript or polypeptide product. Regulatory sequences include, but are not limited to, promoters, promoter control elements, protein binding sequences, 5' and 3' UTRs, transcriptional start site, termination sequence, polyadenylation sequence, introns, certain sequences within a coding sequence, etc.

Related Sequences: "Related sequences" refer to either a polypeptide or a nucleotide sequence that exhibits some degree of sequence similarity with a sequence described by The Reference tables and The Sequence tables.

Scaffold Attachment Region (SAR): As used herein, "scaffold attachment region" is a DNA sequence that anchors chromatin to the nuclear matrix or scaffold to generate loop domains that can have either a transcriptionally active or inactive structure (Spiker and Thompson (1996) Plant Physiol. 110: 15-21).

Sequence-determined DNA fragments (SDFs): "Sequence-determined DNA fragments" as used in the current invention are isolated sequences of genes, fragments of genes, intergenic regions or contiguous DNA from plant genomic DNA or cDNA or RNA the sequence of which has been determined.

Signal Peptide: A "signal peptide" as used in the current invention is an amino acid sequence that targets the protein for secretion, for transport to an intracellular compartment or organelle or for incorporation into a membrane. Signal peptides are indicated in the tables and a more detailed description located below.

Specific Promoter: In the context of the current invention, "specific promoters" refers to a subset of inducible promoters that have a high preference for being induced in a specific tissue or cell and/or at a specific time during development of an organism. By "high preference" is meant at least 3-fold, preferably 5-fold, more preferably at least 10-fold still more preferably at least 20-fold, 50-fold or 100-fold increase in transcription in the desired tissue over the transcription in any other tissue. Typical examples of temporal and/or tissue specific promoters of plant origin that can be used with the polynucleotides of the present invention, are: PTA29, a promoter which is capable of driving gene transcription specifically in tapetum and only during anther development (Koltonow et al., *Plant Cell* 2:1201 (1990); RCc2 and RCc3, promoters that direct root-specific gene transcription in rice (Xu et al., *Plant Mol. Biol.* 27:237 (1995); TobRB27, a root-specific promoter from tobacco (Yamamoto et al., *Plant Cell* 3:371 (1991)). Examples of tissue-specific promoters under developmental control include promoters that initiate transcription only in certain tissues or organs, such as root, ovule, fruit, seeds, or flowers. Other suitable promoters include those from genes encoding storage proteins or the lipid body membrane protein, oleosin. A few root-specific promoters are noted above.

Stringency: "Stringency" as used herein is a function of probe length, probe composition (G + C content), and salt concentration, organic solvent concentration, and temperature of hybridization or wash conditions. Stringency is typically compared by the parameter T_m, which is the temperature at which 50% of the complementary molecules in the hybridization are hybridized, in terms of a temperature differential from T_m. High stringency conditions are those providing a condition of T_m - 5°C to T_m - 10°C. Medium or moderate stringency conditions are those providing T_m - 20°C to T_m - 29°C. Low stringency conditions are those providing a condition of T_m - 40°C to T_m - 48°C. The relationship of hybridization conditions to T_m (in °C) is expressed in the mathematical equation

$$T_m = 81.5 - 16.6(\log_{10}[Na^+]) + 0.41(\%G+C) - (600/N) \quad (1)$$

where N is the length of the probe. This equation works well for probes 14 to 70 nucleotides in length that are identical to the target sequence. The equation below for T_m of DNA-DNA hybrids is useful for probes in the range of 50 to greater than 500 nucleotides, and for conditions that include an organic solvent (formamide).

$$T_m = 81.5+16.6 \log \{[Na^+]/(1+0.7[Na^+])\}+ 0.41(\%G+C)-500/L\ 0.63(\%\text{formamide}) \quad (2)$$

where L is the length of the probe in the hybrid. (P. Tijessen, "Hybridization with Nucleic Acid Probes" in Laboratory Techniques in Biochemistry and Molecular Biology, P.C. vand der Vliet, ed., c. 1993 by Elsevier, Amsterdam.) The T_m of equation (2) is affected by the nature of the hybrid; for DNA-RNA hybrids T_m is 10-15°C higher than calculated, for RNA-RNA hybrids T_m is 20-25°C higher. Because the T_m decreases about 1 °C for each 1% decrease in homology when a long probe is used (Bonner et al., *J. Mol. Biol.* 81:123 (1973)), stringency conditions can be adjusted to favor detection of identical genes or related family members.

Equation (2) is derived assuming equilibrium and therefore, hybridizations according to the present invention are most preferably performed under conditions of probe excess and for sufficient time to achieve equilibrium. The time required to reach equilibrium can be shortened by inclusion of a hybridization accelerator such as dextran sulfate or another high volume polymer in the hybridization buffer.

Stringency can be controlled during the hybridization reaction or after hybridization has occurred by altering the salt and temperature conditions of the wash solutions used. The formulas shown above are equally valid when used to compute the stringency of a wash solution. Preferred wash solution stringencies lie within the ranges stated above; high stringency is 5-8°C below T_m, medium or moderate stringency is 26-29°C below T_m and low stringency is 45-48°C below T_m.

Substantially free of: A composition containing A is "substantially free of " B when at least 85% by weight of the total A+B in the composition is A. Preferably, A comprises at least about 90% by weight of the total of A+B in the composition, more preferably at least about 95% or even 99% by weight. For example, a plant gene or DNA sequence can be considered substantially free of other plant genes or DNA sequences.

Translational start site: In the context of the current invention, a "translational start site" is usually an ATG in the cDNA transcript, more usually the first ATG. A single cDNA, however, may have multiple translational start sites.

Transcription start site: "Transcription start site" is used in the current invention to describe the point at which transcription is initiated. This point is typically located about 25 nucleotides downstream from a TFIID binding site, such as a TATA box. Transcription can initiate at one or more sites within the gene, and a single gene may have multiple transcriptional start sites, some of which may be specific for transcription in a particular cell-type or tissue.

Untranslated region (UTR): A "UTR" is any contiguous series of nucleotide bases that is transcribed, but is not translated. These untranslated regions may be associated with particular functions such as increasing mRNA message stability. Examples of UTRs include, but are not limited to polyadenylation signals, terminations sequences, sequences located between the transcriptional start site and the first exon (5' UTR) and sequences located between the last exon and the end of the mRNA (3' UTR).

Variant: The term "variant" is used herein to denote a polypeptide or protein or polynucleotide molecule that differs from others of its kind in some way. For example, polypeptide and protein variants can consist of changes in amino acid sequence and/or charge and/or post-translational modifications (such as glycosylation, etc).

EXAMPLES

The invention is illustrated by way of the following examples. The invention is not limited by these examples as the scope of the invention is defined solely by the claims following.

EXAMPLE 1: cDNA PREPARATION

A number of the nucleotide sequences disclosed in the Reference and Sequence tables or polynucleotides encoding polypeptides of the Protein Group or Protein Group Matrix tables, herein as representative of the SDFs of the invention can be obtained by sequencing genomic DNA (gDNA) and/or cDNA from corn plants grown from HYBRID SEED # 35A19, purchased from Pioneer Hi-Bred International, Inc., Supply Management, P.O. Box 256, Johnston, Iowa 50131-0256.

A number of the nucleotide sequences disclosed in the Reference and Sequence tables or polynucleotides encoding polypeptides of the Protein Group or Protein Group Matrix tables, herein as representative of the SDFs of the invention can also be obtained by sequencing genomic DNA from *Arabidopsis thaliana*, Wassilewskija ecotype or by sequencing cDNA obtained from mRNA from such plants as described below. This is a true breeding strain. Seeds of the plant are available from the Arabidopsis Biological Resource Center at the Ohio State University, under the accession number CS2360. Seeds of this plant were deposited under the terms and conditions of the Budapest Treaty at the American Type Culture Collection, Manassas, VA on August 31, 1999, and were assigned ATCC No. PTA-595.

Other methods for cloning full-length cDNA are described, for example, by Seki et al., *Plant Journal* 15:707-720 (1998) "High-efficiency cloning of Arabidopsis full-length cDNA by biotinylated Cap trapper"; Maruyama et al., *Gene* 138:171 (1994) "Oligo-capping a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides"; and WO 96/34981.

Tissues were, or each organ was, individually pulverized and frozen in liquid nitrogen. Next, the samples were homogenized in the presence of detergents and then centrifuged. The debris and nuclei were removed from the sample and more detergents were added to the sample. The sample was centrifuged and the debris was removed. Then the sample was applied to a 2M sucrose cushion to isolate polysomes. The RNA was isolated by treatment with detergents and proteinase K followed by ethanol precipitation and centrifugation. The polysomal RNA from the

different tissues was pooled according to the following mass ratios: 15/15/1 for male inflorescences, female inflorescences and root, respectively. The pooled material was then used for cDNA synthesis by the methods described below.

Starting material for cDNA synthesis for the exemplary corn cDNA clones with sequences presented in the Reference and Sequence tables or polynucleotides encoding polypeptides of the Protein Group or Protein Group Matrix tables was poly(A)-containing polysomal mRNAs from inflorescences and root tissues of corn plants grown from HYBRID SEED # 35A19. Male inflorescences and female (pre-and post-fertilization) inflorescences were isolated at various stages of development. Selection for poly(A) containing polysomal RNA was done using oligo d(T) cellulose columns, as described by Cox and Goldberg, "Plant Molecular Biology: A Practical Approach", pp. 1-35, Shaw ed., c. 1988 by IRL, Oxford. The quality and the integrity of the polyA+ RNAs were evaluated.

Starting material for cDNA synthesis for the exemplary *Arabidopsis* cDNA clones with sequences presented in the Reference and Sequence tables or polynucleotides encoding polypeptides of the Protein Group or Protein Group Matrix tables was polysomal RNA isolated from the top-most inflorescence tissues of *Arabidopsis thaliana* Wassilewskija (Ws.) and from roots of *Arabidopsis thaliana* Landsberg erecta (L. er.), also obtained from the Arabidopsis Biological Resource Center. Nine parts inflorescence to every part root was used, as measured by wet mass. Tissue was pulverized and exposed to liquid nitrogen. Next, the sample was homogenized in the presence of detergents and then centrifuged. The debris and nuclei were removed from the sample and more detergents were added to the sample. The sample was centrifuged and the debris was removed and the sample was applied to a 2M sucrose cushion to isolate polysomal RNA. Cox et al., "Plant Molecular Biology: A Practical Approach", pp. 1-35, Shaw ed., c. 1988 by IRL, Oxford. The polysomal RNA was used for cDNA synthesis by the methods described below. Polysomal mRNA was then isolated as described above for corn cDNA. The quality of the RNA was assessed electrophoretically.

Following preparation of the mRNAs from various tissues as described above, selection of mRNA with intact 5' ends and specific attachment of an oligonucleotide tag to the 5' end of such mRNA was performed using either a chemical or enzymatic approach. Both techniques take advantage of the presence of the "cap" structure, which characterizes the 5' end of most intact mRNAs and which comprises a guanosine generally methylated once, at the 7 position.

The chemical modification approach involves the optional elimination of the 2', 3'-cis diol of the 3' terminal ribose, the oxidation of the 2', 3'-cis diol of the ribose linked to the cap of the 5' ends of the mRNAs into a dialdehyde, and the coupling of the such obtained dialdehyde to a derivatized oligonucleotide tag. Further detail regarding the chemical approaches for obtaining mRNAs having intact 5' ends are disclosed in International Application No. WO96/34981 published November 7, 1996.

The enzymatic approach for ligating the oligonucleotide tag to the intact 5' ends of mRNAs involves the removal of the phosphate groups present on the 5' ends of uncapped incomplete mRNAs, the subsequent decapping of mRNAs having intact 5' ends and the ligation of the phosphate present at the 5' end of the decapped mRNA to an oligonucleotide tag. Further detail regarding the enzymatic approaches for obtaining mRNAs having intact 5' ends are disclosed in Dumas Milne Edwards J.B. (Doctoral Thesis of Paris VI University, Le clonage des ADNc complets: difficultés et perspectives nouvelles. Apports pour l'étude de la régulation de l'expression de la tryptophane hydroxylase de rat, 20 Dec. 1993), EP0 625572 and Kato *et al., Gene* 150:243-250 (1994).

In both the chemical and the enzymatic approach, the oligonucleotide tag has a restriction enzyme site (e.g. an EcoRI site) therein to facilitate later cloning procedures. Following attachment of the oligonucleotide tag to the mRNA, the integrity of the mRNA is examined by performing a Northern blot using a probe complementary to the oligonucleotide tag.

For the mRNAs joined to oligonucleotide tags using either the chemical or the enzymatic method, first strand cDNA synthesis is performed using an oligo-dT primer with reverse transcriptase. This oligo-dT primer can contain an internal tag of at least 4 nucleotides, which can be different from one mRNA preparation to another. Methylated dCTP is used for cDNA first strand synthesis to protect the internal EcoRI sites from digestion during subsequent steps. The first strand cDNA is precipitated using isopropanol after removal of RNA by alkaline hydrolysis to eliminate residual primers.

Second strand cDNA synthesis is conducted using a DNA polymerase, such as Klenow fragment and a primer corresponding to the 5' end of the ligated oligonucleotide. The primer is typically 20-25 bases in length. Methylated dCTP is used for second strand synthesis in order to protect internal EcoRI sites in the cDNA from digestion during the cloning process.

Following second strand synthesis, the full-length cDNAs are cloned into a phagemid vector, such as pBlueScript™ (Stratagene). The ends of the full-length cDNAs are blunted with T4 DNA polymerase (Biolabs) and the cDNA is digested with EcoRI. Since methylated dCTP is used during cDNA synthesis, the EcoRI site present in the tag is the only hemi-methylated site; hence the only site susceptible to EcoRI digestion. In some instances, to facilitate subcloning, an Hind III adapter is added to the 3' end of full-length cDNAs.

The full-length cDNAs are then size fractionated using either exclusion chromatography (AcA, Biosepra) or electrophoretic separation which yields 3 to 6 different fractions. The full-length cDNAs are then directionally cloned either into pBlueScript™ using either the EcoRI and SmaI restriction sites or, when the Hind III adapter is present in the full-length cDNAs, the EcoRI and Hind III restriction sites. The ligation mixture is transformed, preferably by electroporation, into bacteria, which are then propagated under appropriate antibiotic selection.

Clones containing the oligonucleotide tag attached to full-length cDNAs are selected as follows.

The plasmid cDNA libraries made as described above are purified (e.g. by a column available from Qiagen). A positive selection of the tagged clones is performed as follows. Briefly, in this selection procedure, the plasmid DNA is converted to single stranded DNA using phage F1 gene II endonuclease in combination with an exonuclease (Chang et al., *Gene* 127:95 (1993)) such as exonuclease III or T7 gene 6 exonuclease. The resulting single stranded DNA is then purified using paramagnetic beads as described by Fry et al., *Biotechniques* 13: 124 (1992). Here the single stranded DNA is hybridized with a biotinylated oligonucleotide having a sequence corresponding to the 3' end of the oligonucleotide tag. Preferably, the primer has a length of 20-25 bases. Clones including a sequence complementary to the biotinylated oligonucleotide are selected by incubation with streptavidin coated magnetic beads followed by magnetic capture. After capture of the positive clones, the plasmid DNA is released from the magnetic beads and converted into double stranded DNA using a DNA polymerase such as ThermoSequenase™ (obtained from Amersham Pharmacia Biotech). Alternatively, protocols such as the Gene Trapper™ kit (Gibco BRL) can be used. The double stranded DNA is then transformed, preferably by electroporation, into bacteria. The percentage of positive clones having the 5' tag oligonucleotide is typically estimated to be between 90 and 98% from dot blot analysis.

Following transformation, the libraries are ordered in microtiter plates and sequenced. The *Arabidopsis* library was deposited at the American Type Culture Collection on January 7, 2000 as "*E-coli* liba 010600" under the accession number PTA-1161.

G. EXAMPLE 2: Southern hybridizations

The SDFs of the invention can be used in Southern hybridizations as described above. The following describes extraction of DNA from nuclei of plant cells, digestion of the nuclear DNA and separation by length, transfer of the separated fragments to membranes, preparation of probes for hybridization, hybridization and detection of the hybridized probe.

The procedures described herein can be used to isolate related polynucleotides or for diagnostic purposes. Moderate stringency hybridization conditions, as defined above, are described in the present example. These conditions result in detection of hybridization between sequences having at least 70% sequence identity. As described above, the hybridization and wash conditions can be changed to reflect the desired percenatge of sequence identity between probe and target sequences that can be detected.

In the following procedure, a probe for hybridization is produced from two PCR reactions using two primers from genomic sequence of *Arabidopsis thaliana.* As described above, the particular template for generating the probe can be any desired template.

The first PCR product is assessed to validate the size of the primer to assure it is of the expected size. Then the product of the first PCR is used as a template, with the same pair of primers used in the first PCR, in a second PCR that produces a labeled product used as the probe.

Fragments detected by hybridization, or other bands of interest, can be isolated from gels used to separate genomic DNA fragments by known methods for further purification and/or characterization.

Buffers for nuclear DNA extraction

1. 10X HB

	1000 ml	
40 mM spermidine	10.2 g	Spermine (Sigma S-2876) and spermidine (Sigma S-2501)

10 mM spermine	3.5 g	Stabilize chromatin and the nuclear membrane
0.1 M EDTA (disodium)	37.2 g	EDTA inhibits nuclease
0.1 M Tris	12.1 g	Buffer
0.8 M KCl	59.6 g	Adjusts ionic strength for stability of nuclei

Adjust pH to 9.5 with 10 N NaOH. It appears that there is a nuclease present in leaves. Use of pH 9.5 appears to inactivate this nuclease.

2. 2 M sucrose (684 g per 1000 ml)
Heat about half the final volume of water to about 50°C. Add the sucrose slowly then bring the mixture to close to final volume; stir constantly until it has dissolved. Bring the solution to volume.

3. Sarkosyl solution (lyses nuclear membranes)

	1000 ml
N-lauroyl sarcosine (Sarkosyl)	20.0 g
0.1 M Tris	12.1 g
0.04 M EDTA (Disodium)	14.9 g

Adjust the pH to 9.5 after all the components are dissolved and bring up to the proper volume.

4. 20% Triton X-100
80 ml Triton X-100
320 ml 1xHB (w/o β-ME and PMSF)

Prepare in advance; Triton takes some time to dissolve

A. Procedure

1. Prepare 1X "H" buffer (keep ice-cold during use)

	1000 ml	
10X HB	100 ml	
2 M sucrose	250 ml	a non-ionic osmoticum
Water	634 ml	

Added just before use:

100 mM PMSF*	10 ml	a protease inhibitor; protects nuclear membrane proteins
ß-mercaptoethanol	1 ml	inactivates nuclease by reducing disulfide bonds

*100 mM PMSF
(phenyl methyl sulfonyl fluoride, Sigma P-7626)
(add 0.0875 g to 5 ml 100% ethanol)

2. Homogenize the tissue in a blender (use 300-400 ml of 1xHB per blender). Be sure that you use 5-10 ml of HB buffer per gram of tissue. Blenders generate heat so be sure to keep the homogenate cold. It is necessary to put the blenders in ice periodically.

3. Add the 20% Triton X-100 (25 ml per liter of homogenate) and gently stir on ice for 20 min. This lyses plastid, but not nuclear, membranes.

4. Filter the tissue suspension through several nylon filters into an ice-cold beaker. The first filtration is through a 250-micron membrane; the second is through an 85-micron membrane; the third is through a 50-micron membrane; and the fourth is through a 20-

micron membrane. Use a large funnel to hold the filters. Filtration can be sped up by gently squeezing the liquid through the filters.

5. Centrifuge the filtrate at 1200 x g for 20 min. at 4°C to pellet the nuclei.

6. Discard the dark green supernatant. The pellet will have several layers to it. One is starch; it is white and gritty. The nuclei are gray and soft. In the early steps, there may be a dark green and somewhat viscous layer of chloroplasts.

Wash the pellets in about 25 ml cold H buffer (with Triton X-100) and resuspend by swirling gently and pipetting. After the pellets are resuspended.

Pellet the nuclei again at 1200 - 1300 x g. Discard the supernatant.

Repeat the wash 3-4 times until the supernatant has changed from a dark green to a pale green. This usually happens after 3 or 4 resuspensions. At this point, the pellet is typically grayish white and very slippery. The Triton X-100 in these repeated steps helps to destroy the chloroplasts and mitochondria that contaminate the prep.

Resuspend the nuclei for a final time in a total of 15 ml of H buffer and transfer the suspension to a sterile 125 ml Erlenmeyer flask.

7. Add 15 ml, dropwise, cold 2% Sarkosyl, 0.1 M Tris, 0.04 M EDTA solution (pH 9.5) while swirling gently. This lyses the nuclei. The solution will become very viscous.

8. Add 30 grams of CsCl and gently swirl at room temperature until the CsCl is in solution. The mixture will be gray, white and viscous.

9. Centrifuge the solution at 11,400 x g at 4°C for at least 30 min. The longer this spin is, the firmer the protein pellicle.

10. The result is typically a clear green supernatant over a white pellet, and (perhaps) under a protein pellicle. Carefully remove the solution under the protein pellicle and above the pellet. Determine the density of the solution by weighing 1 ml of solution and add CsCl if necessary to bring to 1.57 g/ml. The solution contains dissolved solids (sucrose etc) and the refractive index alone will not be an accurate guide to CsCl concentration.

11. Add 20 μl of 10 mg/ml EtBr per ml of solution.

12. Centrifuge at 184,000 x g for 16 to 20 hours in a fixed-angle rotor.

13. Remove the dark red supernatant that is at the top of the tube with a plastic transfer pipette and discard. Carefully remove the DNA band with another transfer pipette. The DNA band is usually visible in room light; otherwise, use a long wave UV light to locate the band.

14. Extract the ethidium bromide with isopropanol saturated with water and salt. Once the solution is clear, extract at least two more times to ensure that all of the EtBr is gone. Be very gentle, as it is very easy to shear the DNA at this step. This extraction may take a while because the DNA solution tends to be very viscous. If the solution is too viscous, dilute it with TE.

15. Dialyze the DNA for at least two days against several changes (at least three times) of TE (10 mM Tris, 1mM EDTA, pH 8) to remove the cesium chloride.

16. Remove the dialyzed DNA from the tubing. If the dialyzed DNA solution contains a lot of debris, centrifuge the DNA solution at least at 2500 x g for 10 min. and carefully transfer the clear supernatant to a new tube. Read the A260 concentration of the DNA.

17. Assess the quality of the DNA by agarose gel electrophoresis (1% agarose gel) of the DNA. Load 50 ng and 100 ng (based on the OD reading) and compare it with known and

good quality DNA. Undigested lambda DNA and a lambda-HindIII-digested DNA are good molecular weight makers.

Protocol for Digestion of Genomic DNA

Protocol:

1. The relative amounts of DNA for different crop plants that provide approximately a balanced number of genome equivalent is given in Table 3. Note that due to the size of the wheat genome, wheat DNA will be underrepresented. Lambda DNA provides a useful control for complete digestion.

2. Precipitate the DNA by adding 3 volumes of 100% ethanol. Incubate at –20°C for at least two hours. Yeast DNA can be purchased and made up at the necessary concentration, therefore no precipitation is necessary for yeast DNA.

3. Centrifuge the solution at 11,400 x g for 20 min. Decant the ethanol carefully (be careful not to disturb the pellet). Be sure that the residual ethanol is completely removed either by vacuum desiccation or by carefully wiping the sides of the tubes with a clean tissue.

4. Resuspend the pellet in an appropriate volume of water. Be sure the pellet is fully resuspended before proceeding to the next step. This may take about 30 min.

5. Add the appropriate volume of 10X reaction buffer provided by the manufacturer of the restriction enzyme to the resuspended DNA followed by the appropriate volume of enzymes. Be sure to mix it properly by slowly swirling the tubes.

6. Set-up the lambda digestion-control for each DNA that you are digesting.

7. Incubate both the experimental and lambda digests overnight at 37°C. Spin down condensation in a microfuge before proceeding.

8. After digestion, add 2 μl of loading dye (typically 0.25% bromophenol blue, 0.25% xylene cyanol in 15% Ficoll or 30% glycerol) to the lambda-control digests and load in 1% TPE-agarose gel (TPE is 90 mM Tris-phosphate, 2 mM EDTA, pH 8). If the lambda DNA in the lambda control digests are completely digested, proceed with the precipitation of the genomic DNA in the digests.

9. Precipitate the digested DNA by adding 3 volumes of 100% ethanol and incubating in –20°C for at least 2 hours (preferably overnight).

EXCEPTION: *Arabidopsis* and yeast DNA are digested in an appropriate volume; they don't have to be precipitated.

10. Resuspend the DNA in an appropriate volume of TE (e.g., 22 μl x 50 blots = 1100 μl) and an appropriate volume of 10X loading dye (e.g., 2.4 μl x 50 blots = 120 μl). Be careful in pipetting the loading dye - it is viscous. Be sure you are pipetting the correct volume.

Table 3

Some guide points in digesting genomic DNA.

Species	Genome Size	Size Relative to Arabidopsis	Genome Equivalent to 2 μg Arabidopsis DNA	Amount of DNA per blot
Arabidopsis	120 Mb	1X	1X	2 μg
Brassica	1,100 Mb	9.2X	0.54X	10 μg
Corn	2,800 Mb	23.3X	0.43X	20 μg

Cotton	2,300 Mb	19.2X	0.52X	20 μg
Oat	11,300 Mb	94X	0.11X	20 μg
Rice	400 Mb	3.3X	0.75X	5 μg
Soybean	1,100 Mb	9.2X	0.54X	10 μg
Sugarbeet	758 Mb	6.3X	0.8X	10 μg
Sweetclover	1,100 Mb	9.2X	0.54X	10 μg
Wheat	16,000 Mb	133X	0.08X	20 μg
Yeast	15 Mb	0.12X	1X	0.25 μg

Protocol for Southern Blot Analysis

The digested DNA samples are electrophoresed in 1% agarose gels in 1x TPE buffer. Low voltage; overnight separations are preferred. The gels are stained with EtBr and photographed.

1. For blotting the gels, first incubate the gel in 0.25 N HCl (with gentle shaking) for about 15 min.

2. Then briefly rinse with water. The DNA is denatured by 2 incubations. Incubate (with shaking) in 0.5 M NaOH in 1.5 M NaCl for 15 min.

3. The gel is then briefly rinsed in water and neutralized by incubating twice (with shaking) in 1.5 M Tris pH 7.5 in 1.5 M NaCl for 15 min.

4. A nylon membrane is prepared by soaking it in water for at least 5 min, then in 6X SSC for at least 15 min. before use. (20x SSC is 175.3 g NaCl, 88.2 g sodium citrate per liter, adjusted to pH 7.0.)

5. The nylon membrane is placed on top of the gel and all bubbles in between are removed. The DNA is blotted from the gel to the membrane using an absorbent medium, such as paper toweling and 6x SCC buffer. After the transfer, the membrane may be lightly brushed with a gloved hand to remove any agarose sticking to the surface.

6. The DNA is then fixed to the membrane by UV crosslinking and baking at 80°C. The membrane is stored at 4°C until use.

B. Protocol for PCR Amplification of Genomic Fragments in Arabidopsis

Amplification procedures:

1. Mix the following in a 0.20 ml PCR tube or 96-well PCR plate:

Volume	Stock	Final Amount or Conc.
0.5 μl	~ 10 ng/μl genomic DNA[1]	5 ng
2.5 μl	10X PCR buffer	20 mM Tris, 50 mM KCl
0.75 μl	50 mM $MgCl_2$	1.5 mM
1 μl	10 pmol/μl Primer 1 (Forward)	10 pmol
1 μl	10 pmol/μl Primer 2 (Reverse)	10 pmol
0.5 μl	5 mM dNTPs	0.1 mM
0.1 μl	5 units/μl Platinum Taq™ (Life Technologies, Gaithersburg, MD) DNA Polymerase	1 units
(to 25 μl)	Water	

[1] Arabidopsis DNA is used in the present experiment, but the procedure is a general one.

2. The template DNA is amplified using a Perkin Elmer 9700 PCR machine:

1) 94°C for 10 min. followed by

2) 5 cycles:	3) 5 cycles:	4) 25 cycles:
94 °C - 30 sec 62 °C - 30 sec 72 °C - 3 min	94 °C - 30 sec 58 °C - 30 sec 72 °C - 3 min	94 °C - 30 sec 53 °C - 30 sec 72 °C - 3 min

5) 72°C for 7 min. Then the reactions are stopped by chilling to 4°C.

The procedure can be adapted to a multi-well format if necessary.

Quantification and Dilution of PCR Products:

1. The product of the PCR is analyzed by electrophoresis in a 1% agarose gel. A linearized plasmid DNA can be used as a quantification standard (usually at 50, 100, 200, and 400 ng). These will be used as references to approximate the amount of PCR products. HindIII-digested Lambda DNA is useful as a molecular weight marker. The gel can be run fairly quickly; e.g., at 100 volts. The standard gel is examined to determine that the size of the PCR products is consistent with the expected size and if there are significant extra bands or smeary products in the PCR reactions.

2. The amounts of PCR products can be estimated on the basis of the plasmid standard.

3. For the small number of reactions that produce extraneous bands, a small amount of DNA from bands with the correct size can be isolated by dipping a sterile 10-μl tip into the band while viewing though a UV Transilluminator. The small amount of agarose gel (with the DNA fragment) is used in the labeling reaction.

C. Protocol for PCR-DIG-Labeling of DNA

Solutions:

Reagents in PCR reactions (diluted PCR products, 10X PCR Buffer, 50 mM $MgCl_2$, 5 U/μl Platinum Taq Polymerase, and the primers)

10X dNTP + DIG-11-dUTP [1:5]: (2 mM dATP, 2 mM dCTP, 2 mM dGTP, 1.65 mM dTTP, 0.35 mM DIG-11-dUTP)

10X dNTP + DIG-11-dUTP [1:10]: (2 mM dATP, 2 mM dCTP, 2 mM dGTP, 1.81 mM dTTP, 0.19 mM DIG-11-dUTP)

10X dNTP + DIG-11-dUTP [1:15]: (2 mM dATP, 2 mM dCTP, 2 mM dGTP, 1.875 mM dTTP, 0.125 mM DIG-11-dUTP)

TE buffer (10 mM Tris, 1 mM EDTA, pH 8)

Maleate buffer: In 700 ml of deionized distilled water, dissolve 11.61 g maleic acid and 8.77 g NaCl. Add NaOH to adjust the pH to 7.5. Bring the volume to 1 L. Stir for 15 min. and sterilize.

10% blocking solution: In 80 ml deionized distilled water, dissolve 1.16g maleic acid. Next, add NaOH to adjust the pH to 7.5. Add 10 g of the blocking reagent powder (Boehringer Mannheim, Indianapolis, IN, Cat. no. 1096176). Heat to 60°C while stirring to dissolve the powder. Adjust the volume to 100 ml with water. Stir and sterilize.

1% blocking solution: Dilute the 10% stock to 1% using the maleate buffer.

Buffer 3 (100 mM Tris, 100 mM NaCl, 50 mM $MgCl_2$, pH9.5). Prepared from autoclaved solutions of 1M Tris pH 9.5, 5 M NaCl, and 1 M $MgCl_2$ in autoclaved distilled water.

Procedure:

1. PCR reactions are performed in 25 μl volumes containing:

PCR buffer	1X
$MgCl_2$	1.5 mM
10X dNTP + DIG-11-dUTP	1X (please see the note below)
Platinum Taq™ Polymerase	1 unit
10 pg probe DNA	
10 pmol primer 1	

Note:	Use for:
10X dNTP + DIG-11-dUTP (1:5)	< 1 kb
10X dNTP + DIG-11-dUTP (1:10)	1 kb to 1.8 kb
10X dNTP + DIG-11-dUTP (1:15)	> 1.8 kb

2. The PCR reaction uses the following amplification cycles:

1) 94°C for 10 min.

2) 5 cycles:	3) 5 cycles:	4) 25 cycles:
95°C - 30 sec 61°C - 1 min 73°C - 5 min	95°C - 30 sec 59°C - 1 min 75°C - 5 min	95°C - 30 sec 51°C - 1 min 73°C - 5 min

5) 72°C for 8 min. The reactions are terminated by chilling to 4°C (hold).

3. The products are analyzed by electrophoresis- in a 1% agarose gel, comparing to an aliquot of the unlabelled probe starting material.

4. The amount of DIG-labeled probe is determined as follows:

Make serial dilutions of the diluted control DNA in dilution buffer (TE: 10 mM Tris and 1 mM EDTA, pH 8) as shown in the following table:

DIG-labeled control DNA starting conc.	Stepwise Dilution	Final Conc. (Dilution Name)
5 ng/μl	1 μl in 49 μl TE	100 pg/μl (A)
100 pg/μl (A)	25 μl in 25 μl TE	50 pg/μl (B)
50 pg/μl (B)	25 μl in 25 μl TE	25 pg/μl (C)
25 pg/μl (C)	20 μl in 30 μl TE	10 pg/μl (D)

a. Serial deletions of a DIG-labeled standard DNA ranging from 100 pg to 10 pg are spotted onto a positively charged nylon membrane, marking the membrane lightly with a pencil to identify each dilution.

b. Serial dilutions (e.g., 1:50, 1:2500, 1:10,000) of the newly labeled DNA probe are spotted.

c. The membrane is fixed by UV crosslinking.

d. The membrane is wetted with a small amount of maleate buffer and then incubated in 1% blocking solution for 15 min at room temp.

e. The labeled DNA is then detected using alkaline phosphatase conjugated anti-DIG antibody (Boehringer Mannheim, Indianapolis, IN, cat. no. 1093274) and an NBT substrate according to the manufacture's instruction.

f. Spot intensities of the control and experimental dilutions are then compared to estimate the concentration of the PCR-DIG-labeled probe.

D. Prehybridization and Hybridization of Southern Blots

Solutions:

100% Formamide purchased from Gibco

20X SSC ($1X = 0.15$ M NaCl, 0.015 M Na_3citrate)

per L: 175 g NaCl

87.5 g $Na_3citrate{\cdot}2H_20$

20% Sarkosyl (N-lauroyl-sarcosine)

20% SDS (sodium dodecyl sulphate)

10% Blocking Reagent: In 80 ml deionized distilled water, dissolve 1.16 g maleic acid. Next, add NaOH to adjust the pH to 7.5. Add 10 g of the blocking reagent powder. Heat to 60°C while stirring to dissolve the powder. Adjust the volume to 100 ml with water. Stir and sterilize.

Prehybridization Mix:

Final Concentration	Components	Volume (per 100 ml)	Stock
50%	Formamide	50 ml	100%
5X	SSC	25 ml	20X
0.1%	Sarkosyl	0.5 ml	20%
0.02%	SDS	0.1 ml	20%
2%	Blocking Reagent	20 ml	10%
	Water	4.4 ml	

General Procedures:

1. Place the blot in a heat-sealable plastic bag and add an appropriate volume of prehybridization solution (30 ml/100cm^2) at room temperature. Seal the bag with a heat sealer, avoiding bubbles as much as possible. Lay down the bags in a large plastic tray (one tray can accommodate at least 4–5 bags). Ensure that the bags are lying flat in the tray so that the prehybridization solution is evenly distributed throughout the bag. Incubate the blot for at least 2 hours with gentle agitation using a waver shaker.

2. Denature DIG-labeled DNA probe by incubating for 10 min. at 98°C using the PCR machine and immediately cool it to 4°C.

3. Add probe to prehybridization solution (25 ng/ml; 30 ml = 750 ng total probe) and mix well but avoid foaming. Bubbles may lead to background.

4. Pour off the prehybridization solution from the hybridization bags and add new prehybridization and probe solution mixture to the bags containing the membrane.

5. Incubate with gentle agitation for at least 16 hours.

6. Proceed to medium stringency post-hybridization wash:

   Three times for 20 min. each with gentle agitation using 1X SSC, 1% SDS at 60°C.

   All wash solutions must be prewarmed to 60°C. Use about 100 ml of wash solution per membrane.

   To avoid background keep the membranes fully submerged to avoid drying in spots; agitate sufficiently to avoid having membranes stick to one another.

7. After the wash, proceed to immunological detection and CSPD development.

E. Procedure for Immunological Detection with CSPD

Solutions:

Buffer 1:	Maleic acid buffer (0.1 M maleic acid, 0.15 M NaCl; adjusted to pH 7.5 with NaoH)
Washing buffer:	Maleic acid buffer with 0.3% (v/v) Tween 20.
Blocking stock solution	10% blocking reagent in buffer 1. Dissolve (10X concentration): blocking reagent powder (Boehringer Mannheim, Indianapolis, IN, cat. no. 1096176) by constantly stirring on a 65°C heating block or heat in a microwave, autoclave and store at 4°C.
Buffer 2 (1X blocking solution):	Dilute the stock solution 1:10 in Buffer 1.
Detection buffer:	0.1 M Tris, 0.1 M NaCl, pH 9.5

Procedure:

1. After the post-hybridization wash the blots are briefly rinsed (1-5 min.) in the maleate washing buffer with gentle shaking.
2. Then the membranes are incubated for 30 min. in Buffer 2 with gentle shaking.
3. Anti-DIG-AP conjugate (Boehringer Mannheim, Indianapolis, IN, cat. no. 1093274) at 75 mU/ml (1:10,000) in Buffer 2 is used for detection. 75 ml of solution can be used for 3 blots.
4. The membrane is incubated for 30 min. in the antibody solution with gentle shaking.
5. The membrane are washed twice in washing buffer with gentle shaking. About 250 mls is used per wash for 3 blots.
6. The blots are equilibrated for 2–5 min in 60 ml detection buffer.

7. Dilute CSPD (1:200) in detection buffer. (This can be prepared ahead of time and stored in the dark at 4°C).

The following steps must be done individually. Bags (one for detection and one for exposure) are generally cut and ready before doing the following steps.

8. The blot is carefully removed from the detection buffer and excess liquid removed without drying the membrane. The blot is immediately placed in a bag and 1.5 ml of CSPD solution is added. The CSPD solution can be spread over the membrane. Bubbles present at the edge and on the surface of the blot are typically removed by gentle rubbing. The membrane is incubated for 5 min. in CSPD solution.
9. Excess liquid is removed and the membrane is blotted briefly (DNA side up) on Whatman 3MM paper. Do not let the membrane dry completely.
10. Seal the damp membrane in a hybridization bag and incubate for 10 min at 37°C to enhance the luminescent reaction.
11. Expose for 2 hours at room temperature to X-ray film. Multiple exposures can be taken. Luminescence continues for at least 24 hours and signal intensity increases during the first hours.

EXAMPLE 3: MICROARRAY EXPERIMENTS AND RESULTS

Example 3: MICROARRAY EXPERIMENTS AND RESULTS

1. Sample Tissue Preparation

(a) Roots

Seeds of *Arabidopsis thaliana* (Ws) were sterilized in full strength bleach for less than 5 min., washed more than 3 times in sterile distilled deionized water and plated on MS agar plates. The plates were placed at 4°C for 3 nights and then placed vertically into a growth chamber having 16 hr light/8 hr dark cycles, 23 °C, 70% relative humidity and ~11,000 LUX. After 2 weeks, the roots were cut from the agar, flash frozen in liquid nitrogen and stored at -80°C.(EXPT REP: 108439 and 108434)

(b) Root Hairless mutants

Plants mutant at the *rhl* gene locus lack root hairs. This mutation is maintained as a heterozygote.

Seeds of *Arabidopsis thaliana* (Landsberg erecta) mutated at the *rhl* gene locus were sterilized using 30% bleach with 1 ul/ml 20% Triton –X 100 and then vernalized at 4°C for 3 days before being plated onto GM agar plates. Plates were placed in growth chamber with 16 hr light/8 hr. dark, 23°C, 14,500-15,900 LUX, and 70% relative humidity for germination and growth.

After 7 days, seedlings were inspected for root hairs using a dissecting microscope. Mutants were harvested and the cotyledons removed so that only root tissue remained. Tissue was then flash frozen in liquid nitrogen and stored at –80C. (EXPT REP: 108433)

Arabidopsis thaliana (Landsberg erecta) seedlings grown and prepared as above were used as controls. (EXPT REP: 108433)

Alternatively, seeds of *Arabidopsis thaliana* (Landsberg erecta), heterozygous for the *rhl1* (root hairless) mutation, were surface-sterilized in 30% bleach containing 0.1% Triton X-100 and further rinsed in sterile water. They were then vernalized at 4° C for 4 days before being plated onto MS agar plates. The plates were maintained in a growth chamber at 24°C with 16 hr light/8 hr dark for germination and growth. After 10 days, seedling roots that expressed the phenotype (i.e. lacking root hairs) were cut below the hypocotyl junction, frozen in liquid nitrogen and stored at –80°C. Those seedlings with the normal root phenotype (heterozygous or wt) were collected as described for the mutant and used as controls.

(c) Rosette Leaves, Stems, and Siliques *Arabidopsis thaliana* (Ws) seed was vernalized at4° C for 3 days before sowing in Metro-mix soil type 350. Flats were placed in a growth chamber having 16 hr light/8 hr dark, 80% relative humidity, 23°C and 13,000 LUX for germination and growth. After 3 weeks, rosette leaves, stems, and siliques (see EXPT REP: 108436, 108437 and 108438) were harvested, flash frozen in liquid nitrogen and stored at -80°C until use. After 4 weeks, siliques (<5mm, 5-10 mm and >10 mm) were harvested, flash frozen in

liquid nitrogen and stored at -80°C until use. 5 week old whole plants (used as controls) were harvested, flash frozen in liquid nitrogen and kept at -80°C until RNA was isolated.

(d) Trichomes

Arabidopsis thaliana (Colombia glabrous) inflorescences were used as a control and CS8143 (hairy inflorescence ecotype) inflorescences, having increased trichomes, were used as the experimental sample.

Approximately 10 µl of each type of seed was sown on a flat of 350 soil (containing 0.03% marathon) and vernalized at 4° C for 3 days. Plants were then grown at room temperature under florescent lighting. Young inflorescences were collected at 30 days for the control plants and 37 days for the experimental plants. Each inflorescence was cut into one-half inch (1/2") pieces, flash frozen in liquid nitrogen and stored at -80°C until RNA was isolated.

(e) Germination

Arabidopsis thaliana seeds (ecotype Ws) were sterilized in bleach and rinsed with sterile water. The seeds were placed in 100mm petri plates containing soaked autoclaved filter paper. Plates were foil-wrapped and left at 4°C for 3 nights to vernalize. After cold treatment, the foil was removed and plates were placed into a growth chamber having 16 hr light/8 hr dark cycles, 23 °C, 70% relative humidity and ~11,000 lux. Seeds were collected 1 d (EXPT REP: 108461), 2 d (EXPT REP: 108462), 3 d (EXPT REP: 108463) and 4 d (EXPT REP: 108464) later, flash frozen in liquid nitrogen and stored at -80°C until RNA was isolated.

(f) Shoot Apical Meristem

Arabidopsis thaliana (Landsberg erecta) plants mutant at the *stm* gene locus lack shoot meristems, produce aerial rosettes, have a reduced number of flowers per inflorescence, as well as a reduced number of petals, stamens and carpels, and is female sterile. This mutation is maintained as a heterozygote.

Seeds of *Arabidopsis thaliana* (Landsberg erecta) mutated at the *stm* locus were sterilized using 30% bleach with 1 ul/ml 20% Triton –X100. The seeds were vernalized at 4°C for 3 days before being plated onto GM agar plates. Half were then put into a 22°C, 24 hr light growth

chamber and half in a 24°C 16 hr light/8 hr dark growth chamber having 14,500-15,900 LUX, and 70% relative humidity for germination and growth.

After 7 days, seedlings were examined for leaf primordia using a dissecting microscope. Presence of leaf primordia indicated a wild type phenotype. Mutants were selected based on lack of leaf primordia. Mutants were then harvested and hypocotyls removed leaving only tissue in the shoot region. Tissue was then flash frozen in liquid nitrogen and stored at –80°C.

Control tissue was isolated from 5 day old Landsberg erecta seedlings grown in the same manner as above. Tissue from the shoot region was harvested in the same manner as the *stm* tissue, but only contained material from the 24°C, 16 hr light/8 hr dark long day cycle growth chamber. (EXPT REP: 108453)

Seeds of maize hybrid 35A (Pioneer) were sown in water-moistened sand in flats (10 rows, 5-6 seed/row) and covered with clear, plastic lids before being placed in a growth chamber having 16 hr light (25°C)/8 hr dark (20°C), 75% relative humidity and 13,000-14,000 LUX. Covered flats were watered every three days for 8 days. Seedlings were carefully removed from the sand and the outer layers of leaf shealth removed. About 2 mm sections were cut and flash frozen in liquid nitrogen prior to storage at –80°C. The tissues above the shoot apices (~1 cm long) were cut, treated as above and used as control tissue.

(g) Abscissic acid (ABA)

Seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown in trays and left at 4°C for 4 days to vernalize. They were then transferred to a growth chamber having grown 16 hr light/8 hr dark, 13,000 LUX, 70% humidity, and 20°C and watered twice a week with 1 L of 1X Hoagland's solution. Approximately 1,000 14 day old plants were spayed with 200-250 mls of 100 μM ABA in a 0.02% solution of the detergent Silwet L-77. Whole seedlings, including roots, were harvested within a 15 to 20 minute time period at 1 hr and 6 hr after treatment, flash-frozen in liquid nitrogen and stored at –80°C.

Seeds of maize hybrid 35A (Pioneer) were sown in water-moistened sand in flats (10 rows, 5-6 seed/row) and covered with clear, plastic lids before being placed in a growth chamber having 16 hr light (25°C)/8 hr dark (20°C), 75% relative humidity and 13,000-14,000 LUX. Covered flats were watered every three days for 7 days. Seedlings were carefully removed from the sand and placed in 1-liter beakers with 100 μM ABA for treatment. Control plants were

treated with water. After 6 hr and 24 hr, aerial and root tissues were separated and flash frozen in liquid nitrogen prior to storage at –80°C.

(h) Auxin Responsive

Seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown in trays and left at 4°C for 4 days to vernalize. They were then transferred to a growth chamber having 16 hr light/8 hr dark, 13,000 LUX, 70% humidity, 20°C and watered twice a week with 1 L of 1X Hoagland's solution (recipe recited in Feldmann et al., (1987) Mol. Gen. Genet. 208: 1-9 and described as complete nutrient solution). Approximately 1,000 14 day old plants were spayed with 200-250 mls of 100 μM NAA in a 0.02% solution of the detergent Silwet L-77. Aerial tissues (everything above the soil line) were harvested within a 15 to 20 minute time period 1 hr and 6 hrs after treatment, flash-frozen in liquid nitrogen and stored at –80°C.

Seeds of maize hybrid 35A (Pioneer) were sown in water-moistened sand in flats (10 rows, 5-6 seed/row) and covered with clear, plastic lids before being placed in a growth chamber having 16 hr light (25°C)/8 hr dark (20°C), 75% relative humidity and 13,000-14,000 LUX. Covered flats were watered every three days for 7 days. Seedlings were carefully removed from the sand and placed in 1-liter beakers with 100 μM NAA for treatment. Control plants were treated with water. After 6 hr and 24 hr, aerial and root tissues were separated and flash frozen in liquid nitrogen prior to storage at –80°C.

(i) Cytokinin

Seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown in trays and left at 4°C for 4 days to vernalize. They were then transferred to a growth chamber having 16 hr light/8 hr dark, 13,000 LUX, 70% humidity, 20°C temperature and watered twice a week with 1 L of 1X Hoagland's solution. Approximately 1,000 14 day old plants were spayed with 200-250 mls of 100 μM BA in a 0.02% solution of the detergent Silwet L-77. Aerial tissues (everything above the soil line) were harvested within a 15 to 20 minute time period 1 hr and 6 hrs after treatment, flash-frozen in liquid nitrogen and stored at –80°C.

Seeds of maize hybrid 35A (Pioneer) were sown in water-moistened sand in flats (10 rows, 5-6 seed/row) and covered with clear, plastic lids before being placed in a growth chamber having 16 hr light (25°C)/8 hr dark (20°C), 75% relative humidity and 13,000-14,000 LUX. Covered flats were watered every three days for 7 days. Seedlings were carefully removed from the sand and placed in 1-liter beakers with 100 μM BA for treatment. Control plants were treated with water. After 6 hr, aerial and root tissues were separated and flash frozen in liquid nitrogen prior to storage at –80°C.

(j) Brassinosteroid Responsive

Two separate experiments were performed, one with epi-brassinolide and one with the brassinosteroid biosynthetic inhibitor brassinazole.

In the epi-brassinolide experiments, seeds of wild-type *Arabidopsis thaliana* (ecotype Wassilewskija) and the brassinosteroid biosynthetic mutant *dwf*4-1 were sown in trays and left at 4°C for 4 days to vernalize. They were then transferred to a growth chamber having 16 hr light/8 hr dark, 11,000 LUX, 70% humidity and 22°C temperature. Four week old plants were spayed with a 1 μM solution of epi-brassinolide and shoot parts (unopened floral primordia and shoot apical meristems) harvested three hours later. Tissue was flash-frozen in liquid nitrogen and stored at –80°C.(EXPT REP 108480)

In the brassinazole experiments, seeds of wild-type *Arabidopsis thaliana* (ecotype Wassilewskija) were grown as described above. Four week old plants were spayed with a 1 μM solution of brassinazole and shoot parts (unopened floral primordia and shoot apical meristems) harvested three hours later. Tissue was flash-frozen in liquid nitrogen and stored at –80°C.(EXPT REP 108481)

In addition to the spray experiments, tissue was prepared from two different mutants; (1) a *dwf*4-1 knock out mutant (EXPT REP: 108478) and (2) a mutant overexpressing the *dwf4*-1 gene (EXPT REP: 108479).

Seeds of wild-type *Arabidopsis thaliana* (ecotype Wassilewskija) and of the *dwf*4-1 knock out and overexpressor mutants were sown in trays and left at 4°C for 4 days to vernalize. They were then transferred to a growth chamber having 16 hr light/8 hr dark, 11,000 LUX, 70% humidity and 22°C temperature. Tissue from shoot parts (unopened floral primordia and shoot apical meristems) was flash-frozen in liquid nitrogen and stored at –80°C.

Another experiment was completed with seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown in trays and left at 4°C for 4 days to vernalize. They were then transferred to a growth chamber. Plants were grown under long-day (16 hr light: 8 hr. dark) conditions, 13,000 LUX light intensity, 70% humidity, 20°C temperature and watered twice a week with 1 L 1X Hoagland's solution(recipe recited in Feldmann et al., (1987) Mol. Gen. Genet. 208: 1-9 and described as complete nutrient solution). Approximately 1,000 14 day old plants were spayed with 200-250 mls of 0.1 μM Epi-Brassinolite in 0.02% solution of the detergent Silwet L-77. At 1 hr. and 6 hrs. after treatment aerial tissues were harvested within a 15 to 20 minute time period and flash-frozen in liquid nitrogen.

Seeds of maize hybrid 35A (Pioneer) were sown in water-moistened sand in flats (10 rows, 5-6 seed/row) and covered with clear, plastic lids before being placed in a growth chamber having 16 hr light (25°C)/8 hr dark (20°C), 75% relative humidity and 13,000-14,000 LUX. Covered flats were watered every three days for 7 days. Seedlings were carefully removed from the sand and placed in 1-liter beakers with 0.1 μM epi-brassinolide for treatment. Control plants were treated with distilled deionized water. After 24 hr, aerial and root tissues were separated and flash frozen in liquid nitrogen prior to storage at –80°C.

(k) Gibberillic Acid

Seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown in trays and left at 4°C for 4 days to vernalize. They were then transferred to a growth chamber having 16 hr light/8 hr. dark, 13,000 LUX, 70% humidity, 20°C and watered twice a week with 1 L of 1X Hoagland's solution. Approximately 1,000 14 day old plants were spayed with 200-250 mls of 100 μM gibberillic acid in a 0.02% solution of the detergent Silwet L-77. At 1 hr. and 6 hrs. after treatment, aerial tissues (everything above the soil line) were harvested within a 15 to 20 minute time period, flash-frozen in liquid nitrogen and stored at –80°C.

Alternatively, seeds of *Arabidopsis thaliana* (ecotype Ws) were sown in Metro-mix soil type 350 and left at 4°C for 3 days to vernalize. They were then transferred to a growth chamber having 16 hr light/8 hr dark, 13,000 LUX, 80% humidity, 20°C temperature and watered every four days with 1.5 L water. 14 days after germination, plants were sprayed with 100 μM

gibberillic acid or with water. Aerial tissues were harvested 1 hr (EXPT REP: 108484), 6 hrs (EXPT REP: 108485), 12 hrs (EXPT REP: 108486), and 24 hrs post-treatment, flash frozen and stored at –80°C.

Seeds of maize hybrid 35A (Pioneer) were sown in water-moistened sand in flats (10 rows, 5-6 seed/row) and covered with clear, plastic lids before being placed in a growth chamber having 16 hr light (25°C)/8 hr dark (20°C), 75% relative humidity and 13,000-14,000 LUX. Covered flats were watered every three days for 7 days. Seedlings were carefully removed from the sand and placed in 1-liter beakers with 100 μM gibberillic acid for treatment. Control plants were treated with water. After 1 hr, 6 hr and 12 hr, aerial and root tissues were separated and flash frozen in liquid nitrogen prior to storage at –80°C.

(l) Nitrogen: High to Low

Wild type *Arabidopsis thaliana* seeds (ecotpye Ws) were surface sterilized with 30% Clorox, 0.1% Triton X-100 for 5 minutes. Seeds were then rinsed with 4-5 exchanges of sterile double distilled deionized water. Seeds were vernalized at 4°C for 2-4 days in darkness. After cold treatment, seeds were plated on modified 1X MS media (without NH_4NO_3 or KNO_3), 0.5% sucrose, 0.5g/L MES pH5.7, 1% phytagar and supplemented with KNO_3 to a final concentration of 60 mM (high nitrate modified 1X MS media). Plates were then grown for 7 days in a Percival growth chamber at 22°C with 16 hr. light/8 hr dark.

Germinated seedlings were then transferred to a sterile flask containing 50 mL of high nitrate modified 1X MS liquid media. Seedlings were grown with mild shaking for 3 additional days at 22°C in 16 hr. light/8 hr dark (in a Percival growth chamber) on the high nitrate modified 1X MS liquid media.

After three days of growth on high nitrate modified 1X MS liquid media, seedlings were transferred either to a new sterile flask containing 50 mL of high nitrate modified 1X MS liquid media or to low nitrate modified 1X MS liquid media (containing 20 □M KNO_3). Seedlings were grown in these media conditions with mild shaking at 22°C in 16 hr light/ 8 hr dark for the appropriate time points and whole seedlings harvested for total RNA isolation via the Trizol method (LifeTech.). The time points used for the microarray experiments were 10 min. (EXPT REP: 108454) and 1 hour (EXPT REP: 108455) time points for both the high and low nitrate modified 1X MS media.

Alternatively, seeds that were surface sterilized in 30% bleach containing 0.1% Triton X-100 and further rinsed in sterile water, were planted on MS agar, (0.5% sucrose) plates containing 50 mM KNO_3 (potassium nitrate). The seedlings were grown under constant light (3500 LUX) at 22°C. After 12 days, seedlings were transferred to MS agar plates containing either 1mM KNO_3 or 50 mM KNO_3. Seedlings transferred to agar plates containing 50 mM KNO_3 were treated as controls in the experiment. Seedlings transferred to plates with 1mM KNO_3 were rinsed thoroughly with sterile MS solution containing 1 mM KNO_3. There were ten plates per transfer. Root tissue was collected and frozen in 15 mL Falcon tubes at various time points which included 1 hour, 2 hours, 3 hours, 4 hours, 6 hours, 9 hours, 12 hours, 16 hours, and 24 hours.

Maize 35A19 Pioneer hybrid seeds were sown on flats containing sand and grown in a Conviron growth chamber at 25°C, 16 hr light/8 hr dark, ~13,000 LUX and 80% relative humidity. Plants were watered every three days with double distilled deionized water. Germinated seedlings are allowed to grow for 10 days and were watered with high nitrate modified 1X MS liquid media (see above). On day 11, young corn seedlings were removed from the sand (with their roots intact) and rinsed briefly in high nitrate modified 1X MS liquid media. The equivalent of half a flat of seedlings were then submerged (up to their roots) in a beaker containing either 500 mL of high or low nitrate modified 1X MS liquid media (see above for details).

At appropriate time points, seedlings were removed from their respective liquid media, the roots separated from the shoots and each tissue type flash frozen in liquid nitrogen and stored at –80°C. This was repeated for each time point. Total RNA was isolated using the Trizol method (see above) with root tissues only.

Corn root tissues isolated at the 4 hr and 16 hr time points were used for the microarray experiments. Both the high and low nitrate modified 1X MS media were used.

<u>(m) Nitrogen: Low to High</u>

Arabidopsis thaliana ecotype Ws seeds were sown on flats containing 4 L of a 1:2 mixture of Grace Zonolite vermiculite and soil. Flats were watered with 3 L of water and vernalized at 4°C for five days. Flats were placed in a Conviron growth chamber having 16 hr light/8 hr dark at 20°C, 80% humidity and 17,450 LUX. Flats were watered with approximately 1.5 L of water every four days. Mature, bolting plants (24 days after germination) were bottom

treated with 2 L of either a control (100 mM mannitol pH 5.5) or an experimental (50 mM ammonium nitrate, pH 5.5) solution. Roots, leaves and siliques were harvested separately 30, 120 and 240 minutes after treatment, flash frozen in liquid nitrogen and stored at –80°C.

Hybrid maize seed (Pioneer hybrid 35A19) were aerated overnight in deionized water. Thirty seeds were plated in each flat, which contained 4 liters of Grace zonolite vermiculite. Two liters of water were bottom fed and flats were kept in a Conviron growth chamber with 16 hr light/8 hr dark at 20°C and 80% humidity. Flats were watered with 1 L of tap water every three days. Five day old seedlings were treated as described above with 2 L of either a control (100 mM mannitol pH 6.5) solution or 1 L of an experimental (50 mM ammonium nitrate, pH 6.8) solution. Fifteen shoots per time point per treatment were harvested 10, 90 and 180 minutes after treatment, flash frozen in liquid nitrogen and stored at –80°C.

Alternatively, seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were left at 4°C for 3 days to vernalize. They were then sown on vermiculite in a growth chamber having 16 hours light/8 hours dark, 12,000-14,000 LUX, 70% humidity, and 20°C. They were bottom-watered with tap water, twice weekly. Twenty-four days old plants were sprayed with either water (control) or 0.6% ammonium nitrate at 4 μL/cm^2 of tray surface. Total shoots and some primary roots were cleaned of vermiculite, flash-frozen in liquid nitrogen and stored at –80°C.

(n) Methyl Jasmonate

Seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown in trays and left at 4°C for 4 days to vernalize before being transferred to a growth chamber having 16 hr light/8 hr. dark, 13,000 LUX, 70% humidity, 20°C temperature and watered twice a week with 1 L of a 1X Hoagland's solution. Approximately 1,000 14 day old plants were spayed with 200-250 mls of 0.001% methyl jasmonate in a 0.02% solution of the detergent Silwet L-77. At 1 hr and 6 hrs after treatment, whole seedlings, including roots, were harvested within a 15 to 20 minute time period, flash-frozen in liquid nitrogen and stored at –80°C.

Seeds of maize hybrid 35A (Pioneer) were sown in water-moistened sand in flats (10 rows, 5-6 seed/row) and covered with clear, plastic lids before being placed in a growth chamber having 16 hr light (25°C)/8 hr dark (20°C), 75% relative humidity and 13,000-14,000 LUX. Covered flats were watered every three days for 7 days. Seedlings were carefully removed from the sand and placed in 1-liter beakers with 0.001% methyl jasmonate for treatment. Control

plants were treated with water. After 24 hr, aerial and root tissues were separated and flash frozen in liquid nitrogen prior to storage at –80°C.

(o) Salicylic Acid

Seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown in trays and left at 4°C for 4 days to vernalize before being transferred to a growth chamber having 16 hr light/8 hr. dark, 13,000 LUX, 70% humidity, 20°C temperature and watered twice a week with 1 L of a 1X Hoagland's solution. Approximately 1,000 14 day old plants were spayed with 200-250 mls of 5 mM salicylic acid (solubilized in 70% ethanol) in a 0.02% solution of the detergent Silwet L-77. At 1 hr and 6 hrs after treatment, whole seedlings, including roots, were harvested within a 15 to 20 minute time period flash-frozen in liquid nitrogen and stored at –80°C.

Alternatively, seeds of wild-type *Arabidopsis thaliana* (ecotype Columbia) and mutant CS3726 were sown in soil type 200 mixed with osmocote fertilizer and Marathon insecticide and left at 4°C for 3 days to vernalize. Flats were incubated at room temperature with continuous light. Sixteen days post germination plants were sprayed with 2 mM SA, 0.02% SilwettL-77 or control solution (0.02% SilwettL-77. Aerial parts or flowers were harvested 1 hr (EXPT REP: 108471 and 108472), 4 hr (EXPT REP: 108469 and 108470), 6 hr (EXPT REP: 108440,) 24 hr (EXPT REP: 108443, 107953 and 107960) and 3 weeks (EXPT REP: 108475, 108476) post-treatment flash frozen and stored at –80°C.

Seeds of maize hybrid 35A (Pioneer) were sown in water-moistened sand in flats (10 rows, 5-6 seed/row) and covered with clear, plastic lids before being placed in a growth chamber having 16 hr light (25°C)/8 hr dark (20°C), 75% relative humidity and 13,000-14,000 LUX. Covered flats were watered every three days for 7 days. Seedlings were carefully removed from the sand and placed in 1-liter beakers with 2 mM SA for treatment. Control plants were treated with water. After 12 hr and 24 hr, aerial and root tissues were separated and flash frozen in liquid nitrogen prior to storage at –80°C.

(p) Wounding

Seeds of *Arabidopsis thaliana* (Wassilewskija) were sown in trays and left at 4°C for three days to vernalize before being transferred to a growth chamber having 16 hr light/8 hr dark, 12,000-14,000 LUX, 70% humidity and 20°C. After 14 days, the leaves were wounded with

forceps. Aerial tissues were harvested 1 hour and 6 hours after wounding. Aerial tissues from unwounded plants served as controls. Tissues were flash-frozen in liquid nitrogen and stored at –80°C.

Seeds of maize hybrid 35A (Pioneer) were sown in water-moistened sand in flats (10 rows, 5-6 seed/row) and covered with clear, plastic lids before being placed in a growth chamber having 16 hr light (25°C)/8 hr dark (20°C), 75% relative humidity and 13,000-14,000 LUX. Covered flats were watered every three days for 7 days. Seedlings were wounded (one leaf nicked by scissors) and placed in 1-liter beakers of water for treatment. Control plants were treated not wounded. After 1 hr and 6 hr aerial and root tissues were separated and flash frozen in liquid nitrogen prior to storage at –80°C.

(q) Drought stress

Seeds of *Arabidopsis thaliana* (Wassilewskija) were sown in pots and left at 4°C for three days to vernalize before being transferred to a growth chamber having 16 hr light/8 hr dark, 150,000-160,000 LUX, 20°C and 70% humidity. After 14 days, aerial tissues were cut and left to dry on 3MM Whatman paper in a petri-plate for 1 hour and 6 hours. Aerial tissues exposed for 1 hour and 6 hours to 3 MM Whatman paper wetted with 1X Hoagland's solution served as controls. Tissues were harvested, flash-frozen in liquid nitrogen and stored at –80°C.

Alternatively, *Arabidopsis thaliana* (Ws) seed was vernalized at 4° C for 3 days before sowing in Metromix soil type 350. Flats were placed in a growth chamber with 23°C, 16 hr light/8 hr. dark, 80% relative humidity, ~13,000 LUX for germination and growth. Plants were watered with 1-1.5 L of water every four days. Watering was stopped 16 days after germination for the treated samples, but continued for the control samples. Rosette leaves and stems (EXPT REP 108477, 108482 and 108483), flowers (see EXPT REP: 108473, 108474) and siliques were harvested 2 d, 3 d, 4 d, 5 d, 6 d and 7 d (EXPT REP: 108473) after watering was stopped. Tissue was flash frozen in liquid nitrogen and kept at -80 °C until RNA was isolated. Flowers and siliques were also harvested on day 8 from plants that had undergone a 7 d drought treatment followed by 1 day of watering (EXPT REP: 108474). Control plants (whole plants) were harvested after 5 weeks, flash frozen in liquid nitrogen and stored as above.

Seeds of maize hybrid 35A (Pioneer) were sown in water-moistened sand in flats (10 rows, 5-6 seed/row) and covered with clear, plastic lids before being placed in a growth chamber

having 16 hr light (25°C)/8 hr dark (20°C), 75% relative humidity and 13,000-14,000 LUX. Covered flats were watered every three days for 7 days. Seedlings were carefully removed from the sand and placed in empty 1-liter beakers at room temperature for treatment. Control plants were placed in water. After 1 hr, 6 hr, 12 hr and 24 hr aerial and root tissues were separated and flash frozen in liquid nitrogen prior to storage at –80°C.

(r) Osmotic stress

Seeds of *Arabidopsis thaliana* (Wassilewskija) were sown in trays and left at 4°C for three days to vernalize before being transferred to a growth chamber having 16 hr light/8 hr dark, 12,000-14,000 LUX, 20°C, and 70% humidity. After 14 days, the aerial tissues were cut and placed on 3 MM Whatman paper in a petri-plate wetted with 20% PEG (polyethylene glycol-M_r 8,000) in 1X Hoagland's solution. Aerial tissues on 3 MM Whatman paper containing 1X Hoagland's solution alone served as the control. Aerial tissues were harvested at 1 hour and 6 hours after treatment, flash-frozen in liquid nitrogen and stored at –80°C.

Seeds of maize hybrid 35A (Pioneer) were sown in water-moistened sand in flats (10 rows, 5-6 seed/row) and covered with clear, plastic lids before being placed in a growth chamber having 16 hr light (25°C)/8 hr dark (20°C), 75% relative humidity and 13,000-14,000 LUX. Covered flats were watered every three days for 7 days. Seedlings were carefully removed from the sand and placed in 1-liter beakers with 20% PEG (polyethylene glycol-M_r 8,000) for treatment. Control plants were treated with water. After 1 hr and 6 hr aerial and root tissues were separated and flash frozen in liquid nitrogen prior to storage at –80°C.

Seeds of maize hybrid 35A (Pioneer) were sown in water-moistened sand in flats (10 rows, 5-6 seed/row) and covered with clear, plastic lids before being placed in a growth chamber having 16 hr light (25°C)/8 hr dark (20°C), 75% relative humidity and 13,000-14,000 LUX. Covered flats were watered every three days for 7 days. Seedlings were carefully removed from the sand and placed in 1-liter beakers with 150mM NaCl for treatment. Control plants were treated with water. After 1 hr, 6hr, and 24 hr aerial and root tissues were separated and flash frozen in liquid nitrogen prior to storage at –80°C.

(s) Heat Shock Treatment

Seeds of *Arabidopsis thaliana* (Wassilewskija) were sown in trays and left at 4°C for three days to vernalize before being transferred to a growth chamber with 16 hr light/8 hr dark, 12,000-14,000 LUX, 70% humidity and 20°C. Fourteen day old plants were transferred to a 42°C growth chamber and aerial tissues were harvested 1 h and 6 h after transfer. Control plants were left at 20°C and aerial tissues were harvested. Tissues were flashfrozen in liquid nitrogen and stored at –80°C.

Seeds of maize hybrid 35A (Pioneer) were sown in water-moistened sand in flats (10 rows, 5-6 seed/row) and covered with clear, plastic lids before being placed in a growth chamber having 16 hr light (25°C)/8 hr dark (20°C), 75% relative humidity and 13,000-14,000 LUX. Covered flats were watered every three days for 7 days. Seedlings were carefully removed from the sand and placed in 1-liter beakers containing 42°C water for treatment. Control plants were treated with water at 25°C. After 1 hr and 6 hr aerial and root tissues were separated and flash frozen in liquid nitrogen prior to storage at –80°C.

(t) Cold Shock Treatment

Seeds of *Arabidopsis thaliana* (Wassilewskija) were sown in trays and left at 4°C for three days to vernalize before being transferred to a growth chamber having 16 hr light/8 hr dark, 12,000-14,000 LUX, 20°C and 70% humidity. Fourteen day old plants were transferred to a 4°C dark growth chamber and aerial tissues were harvested 1 hour and 6 hours later. Control plants were maintained at 20°C and covered with foil to avoid exposure to light. Tissues were flash-frozen in liquid nitrogen and stored at –80°C.

Seeds of maize hybrid 35A (Pioneer) were sown in water-moistened sand in flats (10 rows, 5-6 seed/row) and covered with clear, plastic lids before being placed in a growth chamber having 16 hr light (25°C)/8 hr dark (20°C), 75% relative humidity and 13,000-14,000 LUX. Covered flats were watered every three days for 7 days. Seedlings were carefully removed from the sand and placed in 1-liter beakers containing 4°C water for treatment. Control plants were treated with water at 25°C. After 1 hr and 6 hr aerial and root tissues were separated and flash frozen in liquid nitrogen prior to storage at –80°C.

(u) Oxidative Stress- Hydrogen Peroxide Treatment

Seeds of *Arabidopsis thaliana* (Wassilewskija) were sown in trays and left at 4°C for three days to vernalize. before being transferred to a growth chamber having 16 hr light/8 hr dark, 12,000-14,000 LUX, 20°C and 70% humidity. Fourteen day old plants were sprayed with 5 mM H_2O_2 (hydrogen peroxide) in a 0.02% Silwett L-77 solution. Control plants were sprayed with a 0.02% Silwett L-77 solution. Aerial tissues were harvested 1 hour and 6 hours after spraying, flash-frozen in liquid nitrogen and stored at –80°C.

Seeds of maize hybrid 35A (Pioneer) were sown in water-moistened sand in flats (10 rows, 5-6 seed/row) and covered with clear, plastic lids before being placed in a growth chamber having 16 hr light (25°C)/8 hr dark (20°C), 75% relative humidity and 13,000-14,000 LUX. Covered flats were watered every three days for 7 days. Seedlings were carefully removed from the sand and placed in 1-liter beakers with 5 mM H_2O_2 for treatment. Control plants were treated with water. After 1 hr, 6 hr and 24 hr, aerial and root tissues were separated and flash frozen in liquid nitrogen prior to storage at –80°C.

(v) Nitric Oxide Treatment

Seeds of *Arabidopsis thaliana* (Wassilewskija) were sown in trays and left at 4°C for three days to vernalize before being transferred to a growth chamber having 16 hr light/8 hr dark, 12,000-14,000 LUX, 20°C and 70% humidity. Fourteen day old plants were sprayed with 5 mM sodium nitroprusside in a 0.02% Silwett L-77 solution. Control plants were sprayed with a 0.02% Silwett L-77solution. Aerial tissues were harvested 1 hour and 6 hours after spraying, flash-frozen in liquid nitrogen and stored at –80°C.

Seeds of maize hybrid 35A (Pioneer) were sown in water-moistened sand in flats (10 rows, 5-6 seed/row) and covered with clear, plastic lids before being placed in a growth chamber having 16 hr light (25°C)/8 hr dark (20°C), 75% relative humidity and 13,000-14,000 LUX. Covered flats were watered every three days for 7 days. Seedlings were carefully removed from the sand and placed in 1-liter beakers with 5 mM nitroprusside for treatment. Control plants were treated with water. After 1 hr, 6 hr and 12 hr, aerial and root tissues were separated and flash frozen in liquid nitrogen prior to storage at –80°C.

(w) S4 Immature Buds, Inflorescence Meristem

Seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were then transferred to a growth chamber. Plants were grown under long-day (16 hr light: 8 hr dark) conditions, 7000-8000 LUX light intensity, 70% humidity, and 22°C temperature. Inflorescences containing immature floral buds [stages 1-12; Smyth et al., 1990] as well as the inflorescence meristem were harvested and flash frozen in liquid nitrogen.

(x) S5 Flowers (Opened)

Seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were then transferred to a growth chamber. Plants were grown under long-day (16 hr light: 8 hr dark) conditions, 7000-8000 LUX light intensity, 70% humidity, and 22°C temperature. Mature, unpollinated flowers [stages 12-14; Smyth et al. 1990] were harvested and flash frozen in liquid nitrogen.

(y) S6 Siliques (All Stages)

Seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were then transferred to a growth chamber. Plants were grown under long-day (16 hr light: 8 hr dark) conditions, 7000-8000 LUX light intensity, 70% humidity, and 22°C temperature. Siliques bearing developing seeds containing post fertilization through pre-heart stage [0-72 hours after fertilization (HAF)], heart- through early curled cotyledon stage [72-120 HAF] and late-curled cotyledon stage [>120 HAF] embryos were harvested separately and pooled prior to RNA isolation in a mass ratio of 1:1:1. The tissues were then flash frozen in liquid nitrogen. Description of the stages of Arabidopsis embryogenesis used were reviewed by Bowman (1994).

(z) ARABIDOPSIS ENDOSPERM

***mea/mea* Fruits 0-10 mm**

Seeds of *Arabidopsis thaliana* heterozygous for the *fertilization- independent endosperm1* (*fie1*) [Ohad et al., 1996; ecotype Landsberg *erecta* (L*er*)] were sown in pots and left at 4°C for two to three days to vernalize. Kiyosue et al. (1999) subsequently determined that

fie1 was allelic to the gametophytic maternal effect mutant *medea* (Grossniklaus et al., 1998). Imbibed seeds were then transferred to a growth chamber. Plants were grown under long-day (16 hr light: 8 hr dark) conditions, 7000-8000 LUX light intensity, 70% humidity, and 22°C temperature. 1-2 siliques (fruits) bearing developing seeds just prior to dessication [9 days after flowering (DAF)] were selected from each plant and were hand-dissected to identify wild-type, *mea/+* heterozygotes, and *mea/mea* homozygous mutant plants. At this stage, homozygous *mea/mea* plants produce short siliques that contain >70% aborted seed and can be distinguished from those produced by wild-type (100% viable seed) and *mea/+* heterozygous (50% viable seed) plants (Ohad et al., 1996; Grossniklaus et al., 1998; Kiyosue et al., 1999). Siliques 0-10 mm in length containing developing seeds 0-9 DAF produced by homozygous *mea/mea* plants were harvested and flash frozen in liquid nitrogen.

Pods 0-10 mm (Control Tissue for Sample 70)

Seeds of *Arabidopsis thaliana* heterozygous for the *fertilization- independent endosperm1* (*fie1*) [Ohad et al., 1996; ecotype Landsberg *erecta* (L*er*)] were sown in pots and left at 4°C for two to three days to vernalize. Kiyosue et al. (1999) subsequently determined that *fie1* was allelic to the gametophytic maternal effect mutant *medea* (Grossniklaus et al., 1998). Imbibed seeds were then transferred to a growth chamber. Plants were grown under long-day (16 hr light: 8 hr dark) conditions, 7000-8000 LUX light intensity, 70% humidity, and 22°C temperature. 1-2 siliques (fruits) bearing developing seeds just prior to dessication [9 days after flowering (DAF)] were selected from each plant and were hand-dissected to identify wild-type, *mea/+* heterozygotes, and *mea/mea* homozygous mutant plants. At this stage, homozygous *mea/mea* plants produce short siliques that contain >70% aborted seed and can be distinguished from those produced by wild-type (100% viable seed) and *mea/+* heterozygous (50% viable seed) plants (Ohad et al., 1996; Grossniklaus et al., 1998; Kiyosue et al., 1999). Siliques 0-10 mm in length containing developing seeds 0-9 DAF produced by segregating wild-type plants were opened and the seeds removed. The remaining tissues (pods minus seed) were harvested and flash frozen in liquid nitrogen.

(aa) ARABIDOPSIS SEEDS

Fruits (pod + seed) 0-5 mm

Seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were then transferred to a growth chamber. Plants were grown under long-day (16 hr light: 8 hr dark) conditions, 7000-8000 LUX light intensity, 70% humidity, and 22°C temperature. 3-4 siliques (fruits) bearing developing seeds were selected from at least 3 plants and were hand-dissected to determine what developmental stage(s) were represented by the enclosed embryos. Description of the stages of Arabidopsis embryogenesis used in this determination were summarized by Bowman (1994). Silique lengths were then determined and used as an approximate determinant for embryonic stage. Siliques 0-5 mm in length containing post fertilization through pre-heart stage [0-72 hours after fertilization (HAF)] embryos were harvested and flash frozen in liquid nitrogen.

Fruits(pod + seed) 5-10 mm

Seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were then transferred to a growth chamber. Plants were grown under long-day (16 hr light: 8 hr dark) conditions, 7000-8000 LUX light intensity, 70% humidity, and 22°C temperature. 3-4 siliques (fruits) bearing developing seeds were selected from at least 3 plants and were hand-dissected to determine what developmental stage(s) were represented by the enclosed embryos. Description of the stages of Arabidopsis embryogenesis used in this determination were summarized by Bowman (1994). Silique lengths were then determined and used as an approximate determinant for embryonic stage. Siliques 5-10 mm in length containing heart- through early upturned-U- stage [72-120 hours after fertilization (HAF)] embryos were harvested and flash frozen in liquid nitrogen.

Fruits(pod + seed) >10 mm

Seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were then transferred to a growth chamber. Plants were grown under long-day (16 hr light: 8 hr dark) conditions, 7000-8000 LUX light intensity, 70% humidity, and 22°C temperature. 3-4 siliques (fruits) bearing developing seeds were selected from at least 3 plants and were hand-dissected to determine what developmental stage(s) were

represented by the enclosed embryos. Description of the stages of Arabidopsis embryogenesis used in this determination were summarized by Bowman (1994). Silique lengths were then determined and used as an approximate determinant for embryonic stage. Siliques >10 mm in length containing green, late upturned-U- stage [>120 hours after fertilization (HAF)-9 days after flowering (DAF)] embryos were harvested and flash frozen in liquid nitrogen.

Green Pods 5-10 mm (Control Tissue for Samples 72-74)

Seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were then transferred to a growth chamber. Plants were grown under long-day (16 hr light: 8 hr dark) conditions, 7000-8000 LUX light intensity, 70% humidity, and 22°C temperature. 3-4 siliques (fruits) bearing developing seeds were selected from at least 3 plants and were hand-dissected to determine what developmental stage(s) were represented by the enclosed embryos. Description of the stages of Arabidopsis embryogenesis used in this determination were summarized by Bowman (1994). Silique lengths were then determined and used as an approximate determinant for embryonic stage. Green siliques 5-10 mm in length containing developing seeds 72-120 hours after fertilization (HAF)] were opened and the seeds removed. The remaining tissues (green pods minus seed) were harvested and flash frozen in liquid nitrogen.

Green Seeds from Fruits >10 mm

Seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were then transferred to a growth chamber. Plants were grown under long-day (16 hr light: 8 hr dark) conditions, 7000-8000 LUX light intensity, 70% humidity, and 22°C temperature. 3-4 siliques (fruits) bearing developing seeds were selected from at least 3 plants and were hand-dissected to determine what developmental stage(s) were represented by the enclosed embryos. Description of the stages of Arabidopsis embryogenesis used in this determination were summarized by Bowman (1994). Silique lengths were then determined and used as an approximate determinant for embryonic stage. Green siliques >10

mm in length containing developing seeds up to 9 days after flowering (DAF)] were opened and the seeds removed and harvested and flash frozen in liquid nitrogen.

Brown Seeds from Fruits >10 mm

Seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were then transferred to a growth chamber. Plants were grown under long-day (16 hr light: 8 hr dark) conditions, 7000-8000 LUX light intensity, 70% humidity, and 22°C temperature. 3-4 siliques (fruits) bearing developing seeds were selected from at least 3 plants and were hand-dissected to determine what developmental stage(s) were represented by the enclosed embryos. Description of the stages of Arabidopsis embryogenesis used in this determination were summarized by Bowman (1994). Silique lengths were then determined and used as an approximate determinant for embryonic stage. Yellowing siliques >10 mm in length containing brown, dessicating seeds >11 days after flowering (DAF)] were opened and the seeds removed and harvested and flash frozen in liquid nitrogen.

Green/Brown Seeds from Fruits >10 mm

Seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were then transferred to a growth chamber. Plants were grown under long-day (16 hr light: 8 hr dark) conditions, 7000-8000 LUX light intensity, 70% humidity, and 22°C temperature. 3-4 siliques (fruits) bearing developing seeds were selected from at least 3 plants and were hand-dissected to determine what developmental stage(s) were represented by the enclosed embryos. Description of the stages of Arabidopsis embryogenesis used in this determination were summarized by Bowman (1994). Silique lengths were then determined and used as an approximate determinant for embryonic stage. Green siliques >10 mm in length containing both green and brown seeds >9 days after flowering (DAF)] were opened and the seeds removed and harvested and flash frozen in liquid nitrogen.

Mature Seeds (24 hours after imbibition)

Mature dry seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown onto moistened filter paper and left at 4°C for two to three days to vernalize. Imbibed seeds were then transferred to a growth chamber [16 hr light: 8 hr dark conditions, 7000-8000 LUX light intensity, 70% humidity, and 22°C temperature], the emerging seedlings harvested after 48 hours and flash frozen in liquid nitrogen.

Mature Seeds (Dry)

Seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were then transferred to a growth chamber. Plants were grown under long-day (16 hr light: 8 hr dark) conditions, 7000-8000 LUX light intensity, 70% humidity, and 22°C temperature and taken to maturity. Mature dry seeds are collected, dried for one week at 28°C, and vernalized for one week at 4°C before used as a source of RNA.

Ovules

Seeds of *Arabidopsis thaliana* heterozygous for *pistillata* (*pi*) [ecotype Landsberg *erecta* (L*er*)] were sown in pots and left at 4°C for two to three days to vernalize. They were then transferred to a growth chamber. Plants were grown under long-day (16 hr light: 8 hr dark) conditions, 7000-8000 LUX light intensity, 76% humidity, and 24°C temperature. Inflorescences were harvested from seedlings about 40 days old. The inflorescences were cut into small pieces and incubated in the following enzyme solution (pH 5) at room temperature for 0.5-1 hr.: 0.2% pectolyase Y-23, 0.04% pectinase, 5 mM MES, 3% Sucrose and MS salts (1900 mg/l KNO_3, 1650 mg/l NH_4NO_3, 370 mg/l $MgSO_4 \bullet 7 H_2O$, 170 mg/l KH_2PO_4, 440 mg/l $CaCl_2 \bullet 2 H_2O$, 6.2 mg/l H_2BO_3, 15.6 mg/l $MnSO_4 \bullet 4 H_2O$, 8.6 mg/l $ZnSO_4 \bullet 7 H_2O$, 0.25 mg/l $NaMoO_4 \bullet 2 H_2O$, 0.025 mg/l $CuCO_4 \bullet 5 H_2O$, 0.025 mg/l $CoCl_2 \bullet 6 H_2O$, 0.83 mg/l KI, 27.8 mg/l $FeSO_4 \bullet 7 H_2O$, 37.3 mg/l Disodium EDTA, pH 5.8). At the end of the incubation the mixture of inflorescence material and enzyme solution was passed through a size 60 sieve and then through a sieve with a pore size of 125 μm. Ovules greater than 125 μm in diameter were collected, rinsed twice in B5 liquid medium (2500 mg/l KNO_3, 250 mg/l $MgSO_4 \bullet 7 H_2O$, 150 mg/l NaH2PO4 • H_2O, 150 mg/l $CaCl_2 \bullet 2 H_2O$, 134 mg/l (NH4)2 $CaCl_2 \bullet$ SO_4, 3 mg/l H_2BO_3, 10

mg/l $MnSO_4 \bullet 4 H_2O$, 2 $ZnSO_4 \bullet 7 H_2O$, 0.25 mg/l $NaMoO_4 \bullet 2 H_2O$, 0.025 mg/l $CuCO_4 \bullet 5 H_2O$, 0.025 mg/l $CoCl_2 \bullet 6 H_2O$, 0.75 mg/l KI, 40 mg/l EDTA sodium ferric salt, 20 g/l sucrose, 10 mg/l Thiamine hydrochloride, 1 mg/l Pyridoxine hydrochloride, 1 mg/l Nicotinic acid, 100 mg/l myo-inositol, pH 5.5)), rinsed once in deionized water and flash frozen in liquid nitrogen. The supernatant from the 125 µm sieving was passed through subsequent sieves of 50 µm and 32 µm. The tissue retained in the 32 µm sieve was collected and mRNA prepared for use as a control.

(bb) Herbicide treament

Arabidopsis thaliana (Ws) seeds were sterilized for 5 min. with 30% bleach, 50 µl Triton in a total volume of 50 ml. Seeds were vernalized at 4°C for 3 days before being plated onto GM agar plates at a density of about 144 seeds per plate. Plates were incubated in a Percival growth chamber having 16 hr light/8 hr dark, 80% relative humidity, 22 °C and 11,000 LUX for 14 days.

Plates were sprayed (~0.5 mls/plate) with water, Finale (1.128 g/L), Glean (1.88 g/L), RoundUp (0.01 g/L) or Trimec (0.08 g/L). Tissue was collected and flash frozen in liquid nitrogen at the following time points: 0, 1, 2, 4 (EXPT REP: 107871 (Finale), 107881 (Glean), 107896 (Round-up) and 107886 (Trimec)), 8, 12(EXPT REP: 108467 (Finale), 108468 (Glean), 108465 (Round-up) and 108466, 107891 (Trimec)), and 24 hours. Frozen tissue was stored at -80°C prior to RNA isolation.

(cc) Ap2

Seeds of *Arabidopsis thaliana* (ecotype Landesberg erecta) and floral mutant *apetala2* (Jofuku et al., 1994, Plant Cell 6:1211-1225) were sown in pots and left at 4°C for two to three days to vernalize. They were then transferred to a growth chamber. Plants were grown under long-day (16 hr light, 8 hr dark) conditions 7000-8000 LUX light intensity, 70% humidity and 22 °C temperature. Inflorescences containing immature floral buds (stages 1-7; Bowman, 1994) as wel as the inflorescence meristem were harvested and flashfrozen. Polysomal polyA+ RNA was isolated from tissue according to Cox and Goldberg, 1988).

(dd) Protein Degradation

Arabidopsis thaliana (ecotype Ws) wild-type and 13B12-1 (homozygous) mutant seed were sown in pots containing Metro-mix 350 soil and incubated at 4°C for four days. Vernalized seeds were germinated in the greenhouse (16 hr light/8 hr dark) over a 7 day period. Mutant seedlings were sprayed with 0.02% (active ingredient) Finale to confirm their transgenic standing. Plants were grown until the mutant phenotype (either multiple pistils in a single flower and/or multiple branching per node) was apparent. Young inflorescences immediately forming from the multiple-branched stems were cut and flash frozen in liquid nitrogen. Young inflorescences from wild-type plants grown in parallel and under identical conditions were collected as controls. All collected tissue was stored at -80°C until RNA isolation. (EXPT REP 108451)

(ee) Root tips

Seeds of Arabidopsis thaliana (ecotye Ws) were placed on MS plates and vernalized at 4°C for 3 days before being placed in a 25°C growth chamber having 16 hr light/8 hr dark, 70% relative humidty and about 3 W/m^2. After 6 days, young seedlings were transferred to flasks containing B5 liquid medium, 1% sucrose and 0.05 mg/l indole-3-butyric acid. Flasks were incubated at room temperature with 100 rpm agitation. Media was replaced weekly. After three weeks, roots were harvested and incubated for 1 hr with 2% pectinase, 0.2% cellulase, pH 7 before straining through a #80 (Sigma) sieve. The root body material remaining on the sieve (used as the control) was flash frozen and stored at –80°C until use. The material that passed through the #80 sieve was strained through a #200 (Sigma) sieve and the material remaining on the sieve (root tips) was flash frozen and stored at –80°C until use. Approximately 10 mg of root tips were collected from one flask of root culture.

Seeds of maize hybrid 35A (Pioneer) were sown in water-moistened sand in flats (10 rows, 5-6 seed/row) and covered with clear, plastic lids before being placed in a growth chamber having 16 hr light (25°C)/8 hr dark (20°C), 75% relative humidity and 13,000-14,000 LUX. Covered flats were watered every three days for 8 days. Seedlings were carefully removed from the sand and the root tips (~2 mm long) were removed and flash frozen in liquid nitrogen prior to storage at –80°C. The tissues above the root tips (~1 cm long) were cut, treated as above and used as control tissue.

(ff) rt1

The *rt1* allele is a variation of *rt1 rootless1* and is recessive. Plants displaying the *rt1* phenotype have few or no secondary roots.

Seed from plants segregating for *rt1* were sown on sand and placed in a growth chamber having 16 hr light/8 hr dark, 13,000 LUX, 70% humidity and 20°C temperature. Plants were watered every three days with tap water. Eleven (11) day old seedlings were carefully removed from the sand, keeping the roots intact. *rt1*-type seedlings were separated from their wild-type counterparts and the root tissue isolated. Root tissue from normal seedlings (control) and rt1 mutants were flash frozen in liquid nitrogen and stored at –80°C until use.

(gg) Imbibed seed

Seeds of maize hybrid 35A (Pioneer) were sown in water-moistened sand in covered flats (10 rows, 5-6 seed/row) and covered with clear, plastic lids before being placed in a growth chamber having 16 hr light (25°C)/8 hr dark (20°C), 75% relative humidity and 13,000-14,000 LUX. One day after sowing, whole seeds were flash frozen in liquid nitrogen prior to storage at –80°C. Two days after sowing, embryos and endosperm were isolated and flash frozen in liquid nitrogen prior to storage at –80°C. On days 3-6, aerial tissues, roots and endosperm were isolated and flash frozen in liquid nitrogen prior to storage at –80°C.

(hh) Rough Sheath2-R (rs2-R) Mutants (1400-6/S-17)

This experiment was conducted to identify abnormally expressed genes in the shoot apex of *rough sheath2-R* (*rs2-R*) mutant plants. *rs2* encodes a myb domain DNA binding protein that functions in repression of several shoot apical meristem expressed homeobox genes. Two homeobox gene targets are known for *rs2* repression, *rough sheath1, liguleless 3*. The recessive loss of function phenotype of *rs2*-R homozygous plants is described in Schneeberger et al. 1998 Development **125:** 2857-2865.

The seed stock genetically segregates 1:1 for *rs2-R/rs2-R* : *rs2-R/+*

Preparation of tissue samples: 160 seedlings pooled from 2 and 3 week old plants grown in sand. Growth conditions; Conviron #107 @ 12 hr days/12hr night, 25°C, 75% humidity. Shoot apex was dissected to include leaf three and older. (Pictures available upon request).

1) *rough sheath2-R* homozygous (mutant) shoot apex

2) *rough sheath2-R* heterozygous (wt, control) shoot apex

(ii) Leaf Mutant 3642:

Mutant 3642 is a recessive mutation that causes abnormal leaf development. The leaves of mutant 3642 plants are characterized by leaf twisting and irregular leaf shape. Mutant 3642 plants also exhibit abnormally shaped floral organs which results in reduced fertility.

Seed segregating for the mutant phenotype was sown in Metro-mix 350 soil and grown in a Conviron growth chamber with watering by sub-irrigation twice a week. Environmental conditions were set at 20 degrees Celsius, 70% humidity with an 8 hour day, 16 hour night light regime. Plants were harvested after 4 weeks of growth and the entire aerial portion of the plant was harvested and immediately frozen in liquid nitrogen and stored at -80C. Mutant phenotype plants were harvested separately from normal phenotype plants, which serve as the control tissue.

(jj) Flowers (green, white or buds)

Approximately 10 µl of Arabidopsis thaliana seeds (ecotype Ws) were sown on 350 soil (containing 0.03% marathon) and vernalized at 4C for 3 days. Plants were then grown at room temperature under fluorescent lighting until flowering. Flowers were harvested after 28 days in three different categories. Buds that had not opened at all and were completely green were categorized as "flower buds" (also referred to as green buds by the investigator). Buds that had started to open, with white petals emerging slightly were categorized as "green flowers" (also referred to as white buds by the investigator). Flowers that had opened mostly (with no silique elongation) with white petals completely visible were categorized as "white flowers" (also referred to as open flowers by the investigator). Buds and flowers were harvested with forceps, flash frozen in liquid nitrogen and stored at -80C until RNA was isolated.

2. Microarray Hybridization Procedures

Microarray technology provides the ability to monitor mRNA transcript levels of thousands of genes in a single experiment. These experiments simultaneously hybridize two

differentially labeled fluorescent cDNA pools to glass slides that have been previously spotted with cDNA clones of the same species. Each arrayed cDNA spot will have a corresponding ratio of fluorescence that represents the level of disparity between the respective mRNA species in the two sample pools. Thousands of polynucleotides can be spotted on one slide, and each experiment generates a global expression pattern.

COATING SLIDES

The microarray consists of a chemically coated microscope slide, referred herein as a "chip" with numerous polynucleotide samples arrayed at a high density. The poly-L-lysine coating allows for this spotting at high density by providing a hydrophobic surface, reducing the spreading of spots of DNA solution arrayed on the slides. Glass microscope slides (Gold Seal #3010 manufactured by Gold Seal Products, Portsmouth, New Hampshire, USA) were coated with a 0.1%W/V solution of Poly-L-lysine (Sigma, St. Louis, Missouri) using the following protocol:

Slides were placed in slide racks (Shandon Lipshaw #121). The racks were then put in chambers (Shandon Lipshaw #121).

Cleaning solution was prepared:

70 g NaOH was dissolved in 280 mL ddH2O.

420 mL 95% ethanol was added. The total volume was 700 mL (= 2 X 350 mL); it was stirred until completely mixed.

If the solution remained cloudy, ddH_2O was added until clear.

3. The solution was poured into chambers with slides; the chambers were covered with glass lids. The solution was mixed on an orbital shaker for 2 hr.

4. The racks were quickly transferred to fresh chambers filled with ddH_2O. They were rinsed vigorously by plunging racks up and down.

Rinses were repeated 4X with fresh ddH_2O each time, to remove all traces of NaOH-ethanol.

5. Polylysine solution was prepared:

70 mL poly-L-lysine + 70 mL tissue culture PBS in 560 mL water, using plastic graduated cylinder and beaker.

6. Slides were transferred to polylysine solution and shaken for 1 hr.

7. The rack was transferred to a fresh chambers filled with ddH_2O. It was plunged up and down 5X to rinse.

8. The slides were centrifuged on microtiter plate carriers (paper towels were placed below the rack to absorb liquid) for 5 min. @ 500 rpm. The slide racks were transferred to empty chambers with covers.

9. Slide racks were dried in a 45C oven for 10 min.

10. The slides were stored in a closed plastic slide box.

11. Normally, the surface of lysine coated slides was not very hydrophobic immediately after this process, but became increasingly hydrophobic with storage. A hydrophobic surface helped ensure that spots didn't run together while printing at high densities. After they aged for 10 days to a month the slides were ready to use. However, coated slides that have been sitting around for long periods of time were usually too old to be used. This was because they developed opaque patches, visible when held to the light, and these resulted in high background hybridization from the fluorescent probe.

Alternativey, precoated glass slides were purchased from TeleChem Internation, Inc. (Sunnyvale, CA, 94089; catalog number SMM-25, Superamine substrates).

PCR AMPLIFICATION OF CDNA CLONE INSERTS

Polynucleotides were amplified from Arabidopsis cDNA clones using insert specific probes. The resulting 100uL PCR reactions were purified with Qiaquick 96 PCR purification columns (Qiagen, Valencia, California, USA) and eluted in 30 uL of 5mM Tris. 8.5uL of the elution were mixed with 1.5uL of 20X SSC to give a final spotting solution of DNA in 3X SSC. The concentrations of DNA generated from each clone varied between 10-100 ng/ul, but were usually about 50 ng/ul.

ARRAYING OF PCR PRODUCTS ON GLASS SLIDES

PCR products from cDNA clones were spotted onto the poly-L-Lysine coated glass slides using an arrangement of quill-tip pins (ChipMaker 3 spotting pins; Telechem, International, Inc., Sunnyvale, California, USA) and a robotic arrayer (PixSys 3500, Cartesian Technologies, Irvine, California, USA). Around 0.5 nl of a prepared PCR product was spotted at each location to produce spots with approximately 100um diameters. Spot center-to-center spacing was from 180

um to 210um depending on the array. Printing was conducted in a chamber with relative humidity set at 50%.

Slides containing maize sequences were purchased from Agilent Technology (Palo Alto, CA 94304).

POST-PROCESSING OF SLIDES

After arraying, slides were processed through a series of steps – rehydration, UV cross-linking, blocking and denaturation - required prior to hybridization. Slides were rehydrated by placing them over a beaker of warm water (DNA face down), for 2-3 sec, to distribute the DNA more evenly within the spots, and then snap dried on a hot plate (DNA side, face up). The DNA was then cross-linked to the slides by UV irradiation (60-65mJ; 2400 Stratalinker, Stratagene, La Jolla, California, USA).

Following this a blocking step was performed to modify remaining free lysine groups, and hence minimize their ability to bind labeled probe DNA. To achieve this the arrays were placed in a slide rack. An empty slide chamber was left ready on an orbital shaker. The rack was bent slightly inwards in the middle, to ensure the slides would not run into each other while shaking. The blocking solution was prepared as follows:

3x 350-ml glass chambers (with metal tops) were set to one side, and a large round Pyrex dish with dH_2O was placed ready in the microwave. At this time, 15ml sodium borate was prepared in a 50 ml conical tube.

6-g succinic anhydride was dissolved in approx. 325-350 mL 1-methyl-2-pyrrolidinone. Rapid addition of reagent was crucial.

a. Immediately after the last flake of the succinic anhydride dissolved, the 15-mL sodium borate was added.

b. Immediately after the sodium borate solution mixed in, the solution was poured into an empty slide chamber.

c. The slide rack was plunged rapidly and evenly in the solution. It was vigorously shaken up and down for a few seconds, making sure slides never left the solution.

d. It was mixed on an orbital shaker for 15-20 min. Meanwhile, the water in the Pyrex dish (enough to cover slide rack) was heated to boiling.

Following this, the slide rack was gently plunge in the 95C water (just stopped boiling) for 2 min. Then the slide rack was plunged 5X in 95% ethanol. The slides and rack were centrifuged for 5 min. @ 500 rpm. The slides were loaded quickly and evenly onto the carriers to avoid streaking. The arrays were used immediately or store in slide box.

The Hybridization process began with the isolation of mRNA from the two tissues (see *"Isolation of total RNA" and "Isolation of mRNA"*, below) in question followed by their conversion to single stranded cDNA (see *"Generation of probes for hybridization"*, below). The cDNA from each tissue was independently labeled with a different fluorescent dye and then both samples were pooled together. This final differentially labeled cDNA pool was then placed on a processed microarray and allowed to hybridize (see *"Hybridization and wash conditions"*, below).

ISOLATION OF TOTAL RNA

Approximately 1 g of plant tissue was ground in liquid nitrogen to a fine powder and transferred into a 50-ml centrifuge tube containing 10 ml of Trizol reagent. The tube was vigorously vortexed for 1 min and then incubated at room temperature for 10-20 min. on an orbital shaker at 220 rpm. Two ml of chloroform was added to the tube and the solution vortexed vigorously for at least 30-sec before again incubating at room temperature with shaking. The sample was then centrifuged at 12,000 X g (10,000 rpm) for 15-20 min at 4^oC. The aqueous layer was removed and mixed by inversion with 2.5 ml of 1.2 M NaCl/0.8 M Sodium Citrate and 2.5 ml of isopropyl alcohol added. After a 10 min. incubation at room temperature, the sample was centrifuged at 12,000 X g (10,000 rpm) for 15 min at 4^oC. The pellet was washed with 70% ethanol, re-centrifuged at 8,000 rpm for 5 min and then air dried at room temperature for 10 min. The resulting total RNA was dissolved in either TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) or DEPC (diethylpyrocarbonate) treated deionized water (RNAse-free water). For subsequent isolation of mRNA using the Qiagen kit, the total RNA pellet was dissolved in RNAse-free water.

ISOLATION OF mRNA

mRNA was isolated using the Qiagen Oligotex mRNA Spin-Column protocol (Qiagen, Valencia,California). Briefly, 500 µl OBB buffer (20 mM Tris-Cl, pH 7.5, 1 M NaCl, 2 mM EDTA, 0.2% SDS) was added to 500 µl of total RNA (0.5 – 0.75 mg) and mixed thoroughly. The sample was first incubated at 70°C for 3 min, then at room temperature for 10 minutes and finally centrifuged for 2 min at 14,000 – 18,000 X g. The pellet was resuspended in 400 µl OW2 buffer (10 mM Tris-Cl, pH 7.5, 150 mM NaCl, 1 mM EDTA) by vortexing, the resulting solution placed on a small spin column in a 1.5 ml RNase-free microcentrifuge tube and centrifuged for 1 min at 14,000 – 18,000 X g. The spin column was transferred to a new 1.5 ml RNase-free microcentrifuge tube and washed with 400 µl of OW2 buffer. To release the isolated mRNA from the resin, the spin column was again transferred to a new RNase-free 1.5 ml microcentrifuge tube, 20-100 µl 70°C OEB buffer (5 mM Tris-Cl, pH 7.5) added and the resin resuspended in the resulting solution via pipeting. The mRNA solution was collected after centrifuging for 1 min at 14,000 – 18,000 X g.

Alternatively, mRNA was isolated using the Stratagene Poly(A) Quik mRNA Isolation Kit (Startagene, La Jolla, California). Here, up to 0.5 mg of total RNA (maximum volume of 1 ml) was incubated at 65°C for 5 minutes, snap cooled on ice and 0.1X volumes of 10X sample buffer (10mM Tris-HCl (pH 7.5), 1 mM EDTA (pH 8.0) 5 M NaCl) added. The RNA sample was applied to a prepared push column and passed through the column at a rate of ~1 drop every 2 sec. The solution collected was reapplied to the column and collected as above. 200 µl of high salt buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.5 NaCl) was applied to the column and passed through the column at a rate of ~1 drop every 2 sec. This step was repeated and followed by three low salt buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.1 M NaCl) washes preformed in a similar manner. mRNA was eluted by applying to the column four separate 200 µl aliquots of elution buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA) preheated to 65°C. Here, the elution buffer was passed through the column at a rate of 1 drop/sec. The resulting mRNA solution was precipitated by adding 0.1X volumes of 10X sample buffer, 2,5 volumes of ice-cold 100% ethanol, incubating overnight at –20°C and centrifuging at 14,000-18,000 X g for 20-30 min at 4°C. The pellet was washed with 70% ethanol and air dried for 10 min. at room temperature before resuspension in RNase-free deionized water.

PREPARATION OF YEAST CONTROLS

Plasmid DNA was isolated from the following yeast clones using Qiagen filtered maxiprep kits (Qiagen, Valencia, California): YAL022c(Fun26), YAL031c(Fun21), YBR032w, YDL131w, YDL182w, YDL194w, YDL196w, YDR050c and YDR116c. Plasmid DNA was linearized with either *Bsr*BI (YAL022c(Fun26), YAL031c(Fun21), YDL131w, YDL182w, YDL194w, YDL196w, YDR050c) or *Afl*III (YBR032w, YDR116c) and isolated.

In Vitro Transcription of Yeast Clones

The following solution was incubated at 37°C for 2 hours: 17 μl of isolated yeast insert DNA (1 μg), 20 μl 5X buffer, 10 μl 100 mM DTT, 2.5 μl (100 U) RNasin, 20 μl 2.5 mM (ea.) rNTPs, 2.7 μl (40U) SP6 polymerase and 27.8 μl RNase-free deionized water. 2 μl (2 U) Ampli DNase I was added and the incubation continued for another 15 min. 10 μl 5M NH_4OAC and 100 μl phenol:chloroform:isoamyl alcohol (25:24:1) were added, the solution vortexed and then centrifuged to separate the phases. To precipitate the RNA, 250 μl ethanol was added and the solution incubated at -20°C for at least one hour. The sample was then centrifuged for 20 min at 4°C at 14,000-18,000 X g, the pellet washed with 500 μl of 70% ethanol, air dried at room temperature for 10 min and resuspended in 100 μl of RNase-free deionized water. The precipitation procedure was then repeated.

Alternatively, after the two-hour incubation, the solution was extracted with phenol/chloroform once before adding 0.1 volume 3M sodium acetate and 2.5 volumes of 100% ethanol. The solution was centrifuged at 15,000rpm, 4°C for 20 minutes and the pellet resuspended in RNase-free deionized water. The DNase I treatment was carried out at 37°C for 30 minutes using 2 U of Ampli DNase I in the following reaction condition: 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$. The DNase I reaction was then stopped with the addition of NH_4OAC and phenol:chloroform:isoamyl alcohol (25:24:1), and RNA isolated as described above.

0.15-2.5 ng of the *in vitro* transcript RNA from each yeast clone were added to each plant mRNA sample prior to labeling to serve as positive (internal) probe controls.

GENERATION OF PROBES FOR HYBRIDIZATION

Generation of labeled probes for hybridization from first-strand cDNA

Hybridization probes were generated from isolated mRNA using an Atlas™ Glass Fluorescent Labeling Kit (Clontech Laboratories, Inc., Palo Alto, California, USA). This entails

a two step labeling procedure that first incorporates primary aliphatic amino groups during cDNA synthesis and then couples fluorescent dye to the cDNA by reaction with the amino functional groups. Briefly, 5 µg of oligo(dT)$_{18}$ primer d(TTTTTTTTTTTTTTTTTTV) was mixed with Poly A+ mRNA (1.5 - 2 µg mRNA isolated using the Qiagen Oligotex mRNA Spin-Column protocol or-the Stratagene Poly(A) Quik mRNA Isolation protocol (Stratagene, La Jolla, California, USA)) in a total volume of 25 µl. The sample was incubated in a thermocycler at 70°C for 5 min, cooled to 48°C and 10 µl of 5X cDNA Synthesis Buffer (kit supplied), 5 µl 10X dNTP mix (dATP, dCTP, dGTP, dTTP and aminoallyl-dUTP; kit supplied), 7.5 µl deionized water and 2.5 µl MMLV Reverse Transcriptase (500U) added. The reaction was then incubated at 48°C for 30 minutes, followed by 1hr incubation at 42°C. At the end of the incubation the reaction was heated to 70°C for 10 min, cooled to 37°C and 0.5 µl (5 U) RNase H added, before incubating for 15 min at 37°C. The solution was vortexed for 1 min after the addition of 0.5 µl 0.5 M EDTA and 5 µl of QuickClean Resin (kit supplied) then centrifuged at 14,000-18,000 X g for 1 min. After removing the supernatant to a 0.45 µm spin filter (kit supplied), the sample was again centrifuged at 14,000-18,000 X g for 1 min, and 5.5 µl 3 M sodium acetate and 137.5 µl of 100% ethanol added to the sample before incubating at –20°C for at least 1 hr. The sample was then centrifuged at 14,000-18,000 X g at 4°C for 20 min, the resulting pellet washed with 500 µl 70% ethanol, air-dried at room temperature for 10 min and resuspended in 10 µl of 2X fluorescent labeling buffer (kit provided). 10 µl each of the fluorescent dyes Cy3 and Cy5 (Amersham Pharmacia (Piscataway, New Jersey, USA); prepared according to Atlas™ kit directions of Clontech) were added and the sample incubated in the dark at room temperature for 30 min.

The fluorescently labeled first strand cDNA was precipitated by adding 2 µl 3M sodium acetate and 50 µl 100% ethanol, incubated at –20°C for at least 2 hrs, centrifuged at 14,000-18,000 X g for 20 min, washed with 70% ethanol, air-dried for 10 min and dissolved in 100 µl of water.

Alternatively, 3-4 µg mRNA, 2.5 (~8.9 ng of in vitro translated mRNA) µl yeast control and 3 µg oligo dTV (TTTTTTTTTTTTTTTTTT(A/C/G); Sequence ID No.: X) were mixed in a total volume of 24.7 µl. The sample was incubated in a thermocycler at 70°C for 10 min. before chilling on ice. To this, 8 µl of 5X first strand buffer (<u>SuperScript II RNase H- Reverse</u>

Transcriptase kit from Invitrogen (Carlsbad, California 92008); cat no. 18064022), 0.8 °C of aa-dUTP/dNTP mix (50X; 25mM dATP, 25mM dGTP, 25mM dCTP, 15mM dTTP, 10mM aminoallyl-dUTP), 4 µl of 0.1 M DTT and 2.5 µl (500 units) of Superscript R.T.II enzyme (Stratagene) were added. The sample was incubated at 42°C for 2 hours before a mixture of 10 °C of 1M NaOH and 10°C of 0.5 M EDTA were added. After a 15 minute incubation at 65°C, 25 µl of 1 M Tris pH 7.4 was added. This was mixed with 450 µl of water in a Microcon 30 column before centrifugation at 11,000 X g for 12 min. The column was washed twice with 450 µl (centrifugation at 11,000 g, 12 min.) before eluting the sample by inverting the Microcon column and centrifuging at 11,000 X g for 20 seconds. Sample was dehydrated by centrifugation under vacuum and stored at -20°C.

Each reaction pellet was disolved in 9 µl of 0.1 M carbonate buffer (0.1M sodium carbonate and sodium bicarbonate, pH=8.5-9) and 4.5 µl of this placed in two microfuge tubes. 4.5 µl of each dye (in DMSO) were added and the mixture incubated in the dark for 1 hour. 4.5 µl of 4 M hydroxylamine was added and again incubated in the dark for 15 minutes.

Regardless of the method used for probe generation, the probe was purified using a Qiagen PCR cleanup kit (Qiagen, Valencia, California, USA), and eluted with 100 ul EB (kit provided). The sample was loaded on a Microcon YM-30 (Millipore, Bedford, Massachusetts, USA) spin column and concentrated to 4-5 ul in volume. Probes for the maize microarrays were generated using the Fluorescent Linear Amplification Kit (cat. No. G2556A) from Agilent Technologies (Palo Alto, CA).

HYBRIDIZATION AND WASH CONDITIONS

The following Hybridization and Washing Condition were developed:

Hybridization Conditions:

Labeled probe was heated at 95°C for 3 min and chilled on ice. Then 25 □L of the hybridization buffer which was warmed at 42C was added to the probe, mixing by pipetting, to give a final concentration of:

50% formamide

4x SSC

0.03% SDS

5x Denhardt's solution

0.1 μg/ml single-stranded salmon sperm DNA

The probe was kept at 42C. Prior to the hybridization, the probe was heated for 1 more min., added to the array, and then covered with a glass cover slip. Slides were placed in hybridization chambers (Telechem, Sunnyvale, California) and incubated at 42^0C overnight.

Washing Conditions:

A. Slides were washed in 1x SSC + 0.03% SDS solution at room temperature for 5 minutes,

B. Slides were washed in 0.2x SSC at room temperature for 5 minutes,

C. Slides were washed in 0.05x SSC at room temperature for 5 minutes.

After A, B, and C, slides were spun at 800 x g for 2 min. to dry. They were then scanned.

Maize microarrays were hybridized according to the instructions included Fluorescent Linear Amplification Kit (cat. No. G2556A) from Agilent Technologies (Palo Alto, CA).

SCANNING OF SLIDES

The chips were scanned using a ScanArray 3000 or 5000 (General Scanning, Watertown, Massachusetts, USA). The chips were scanned at 543 and 633nm, at 10 um resolution to measure the intensity of the two fluorescent dyes incorporated into the samples hybridized to the chips.

DATA EXTRACTION AND ANALYSIS

The images generated by scanning slides consisted of two 16-bit TIFF images representing the fluorescent emissions of the two samples at each arrayed spot. These images were then quantified and processed for expression analysis using the data extraction software Imagene ™ (Biodiscovery, Los Angeles, California, USA). Imagene output was subsequently analyzed using the analysis program Genespring ™ (Silicon Genetics, San Carlos, California,

USA). In Genespring, the data was imported using median pixel intensity measurements derived from Imagene output. Background subtraction, ratio calculation and normalization were all conducted in Genespring. Normalization was achieved by breaking the data in to 32 groups, each of which represented one of the 32 pin printing regions on the microarray. Groups consist of 360 to 550 spots. Each group was independently normalized by setting the median of ratios to one and multiplying ratios by the appropriate factor.

EXAMPLE 4: AFLP EXPERIMENTS AND RESULTS

Production of Samples

mRNA was prepared from 27 plant tissues. Based on preliminary cDNA-AFLP analysis with a few primer combinations, 11 plant tissues and/or pooled samples were selected. Samples were selected to give the greatest representation of unique band upon electrophoresis. The final 11 samples or pooled samples used in the cDNA-AFLP analysis were:

S1	Dark adapted seedlings
S2	Roots/Etiolated Seedlings
S3	Mature leaves, soil grown
S4	Immature buds, inflorescence meristem
S5	Flowers opened
S6	Siliques, all stages
S7	Senescing leaves (just beginning to yellow)
S8	Callus Inducing medium
	Callus shoot induction
	Callus root induction
S9	Wounding
	Methyl-jasmonate-treated
S10	Oxidative stress
	Drought stress
	Oxygen Stress-flooding
S11	Heat treated light grown seedling

Cold treated light grown seedlings

cDNA from each of the 11 samples was digested with two restriction endonucleases, namely *Taq*I and *Mse*I. *Taq*I and *Mse*I adapters were then ligated to the restriction enzyme fragments. Using primers to these adapters that were specific in sequence (i.e. without extensions), the restriction fragments were subjected to cycles of non-radioactive pre-amplification.

Selective PCR

In order to limit the number of fragments or bands on each lane of the AFLP gel, fragments were subjected to another round of selective radioactive polymerase chain amplification. The *Taq*I primers used in this amplification were 5'-labelled with P^{33}. For these amplifications, the *Taq*I primers had two extra nucleotides at their 3' end and the *Mse*I primers had three extra nucleotides at their 3' end. This resulted in 16 primer designs for the *Taq*I primer and 64 primer designs for the *Mse*I primer. Altogether, this gave rise to a total of 1024 primer designs. Fragments generated in this selective amplification protocol were run with labeled molecular weight markers on polyacrylamide gels to separate fragments in the size range of 100 – 600 nucleotides.

Following gel electrophoresis, profiles were analyzed with a phosphoimager. From these images, electronic files, giving the mobilities of all bands on the gels and their intensities in each of the samples, were compiled.

All unique bands were cut out of the gels. The gel pieces were placed in 96 well plates for elution and their plate designation was linked to their electrophoretic mobilities recorded in the electronic files. The eluted fragments were then subjected to another round of amplification, this time using reamplification primers (see below). After amplification, DNA fragments were sequenced.

A computer database was established linking the mobilities of all the bands observed on the cDNA-AFLP gels with the sequence of the correspondingly isolated fragment. The sequence allowed for identification of the gene from which the cDNA-AFLP fragment was derived, allowing for a linkage of band mobility with the transcript of a specific gene. Also linked to the

band mobilities were their intensities recorded for each of the eleven samples used in constructing the database.

This cDNA-AFLP analysis with *Taq*I/*Mse*I and 1024 primer combinations was repeated using the enzymes *Nla*III in place of *Taq*I, and *Csp*6I in place of *Mse*I.

Using the Database for the Transcript Profiling of Experimental Samples

Experimental Samples were subjected to cDNA-AFLP as described above, resulting in electronic files recording band mobilities and intensities. Through use of the database established above, band mobilities could be linked to specific cDNAs, and therefore genes. Furthermore, the linkage with the intensities in the respective samples allowed for the quantification of specific cDNAs in these samples, and thus the relative concentration of specific transcripts in the samples, indicating the level to which specific genes were expressed.

Reamplification primers

99G24

CGCCAGGGTTTTCCCAGTCACGAC|*ACGACTCACT*|gatgagtcctgagtaa|

M13 forward +10 MseI+0

99G20

AGCGGATAACAATTTCACACAGGA|*CACACTGGTA*|tagactgcgtaccga|

M13 reverse +10 TaqI +0

Purification of the Reamplifiction reaction before sequencing

5 μl reamplification reaction

0,25 μl 10xPCR buffer

0,33 μl Shrimp Alkaline Phosphatase (Amersham Life Science)

0,033 μl Exonuclease I (USB)

0,297 μl SAP dilution buffer

1,59 μl MQ

7.5 μl total

30' 37°C

10' 80°C

4°C

Sample Preparation

S1: Dark adapted seedlings: Seeds of *Arabidopsis thaliana* (wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were transferred to a growth chamber after three days. The intensity of light in the growth chamber was 7000-8000 LUX, temperature was 22°C, with 16 h light and 8 h dark. After 8 days, the seedlings were foil-wrapped and harvested after two days.

S2: Roots/Etiolated seedlings: Seeds of *Arabidopsis thaliana* (wassilewskija) were germinated on solid germination media (1X MS salts, 1X MS vitamins, 20g/L sucrose, 50 mg/L MES pH 5.8) in the dark. Tissues were harvested 14 days later.

S3: Mature leaves, soil grown: Seeds of *Arabidopsis thaliana* (wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were transferred to a growth chamber after three days. The intensity of light in the growth chamber was 7000-8000 LUX, temperature was 22°C, with 16 h light and 8 h dark. Leaves were harvested 17 days later from plants that had not yet bolted.

S4: Immature buds, inflorescence meristem: Seeds of *Arabidopsis thaliana* (wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were transferred to a growth chamber after three days. The intensity of light in the growth chamber was 7000-8000 LUX, temperature was 22°C, with 16 h light and 8 h dark.

S5: Flowers, opened: Seeds of *Arabidopsis thaliana* (wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were transferred to a growth chamber after three days. The intensity of light in the growth chamber was 7000-8000 LUX, temperature was 22°C, with 16 h light and 8 h dark.

S6: Siliques, all stages: Seeds of *Arabidopsis thaliana* (wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were transferred to a growth chamber after three days. The intensity of light in the growth chamber was 7000-8000 LUX, temperature was 22°C, with 16 h light and 8 h dark.

S7: Senescing leaves (just beginning to yellow): Seeds of *Arabidopsis thaliana* (wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were transferred to a growth chamber after three days. The intensity of light in the growth chamber was 7000-8000 LUX, temperature was 22°C, with 16 h light and 8 h dark. When the plant had leaves that were less than 50% yellow, the leaves that were just beginning to yellow were harvested.

S8:

Callus Inducing Medium: Seeds of Arabidopsis thaliana (wassilewskija) were surface sterilized (1 min-75% Ethanol, 6 min-bleach 100% + Tween 20, rinse) and incubated on MS medium containing 2,4-Dichlorophenoxyacetic acid (2,4-D) 1 mg/l and Kinetin 1 mg/l in the dark for 3 weeks to generate primary callus.
Hypocotyls and roots of the seedling were swollen after a week after incubation in this callus induction medium and subsequently callus was initiated from these swollen areas.

Callus shoot induction: Primary calluses were transferred to the fresh callus induction medium for another 2 weeks growth to generate secondary callus. Secondary callus were transferred to shoot induction medium containing MS basal medium and Benzyladenine (BA) 2 mg/l and Naphthaleneacetic acid (NAA)).1 mg/l for 2 weeks growth in the light before it was harvested and frozen and sent to Keygene. Many shoot meristems were observed under the microscope.

Callus root induction: Secondary calluses were transferred to root induction medium containing MS basal medium, sucrose 1% and Indolebutyric acid (IBA) 0.05 mg/l in the dark. Many root primordia were observed under microscope after 10 days in the root induction medium. Those callus tissue were harvested and frozen and sent to Keygene.

S9:

Wounding: Seeds of *Arabidopsis thaliana* (wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were transferred to a growth chamber after three days. The intensity of light in the growth chamber was 7000-8000 LUX, temperature was 22°C, with 16 h light and 8 h dark. After 20 days, leaves of plants were wounded with pliers. Wounded leaves were harvested I hour and 4 hours after wounding.

Methyl jasmonate treatment: Seeds of *Arabidopsis thaliana* (wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were transferred to a growth chamber after three days. The intensity of light in the growth chamber was 7000-8000 LUX, temperature was 22°C, with 16 h light and 8 h dark. After 13 days, plants were sprayed with 0.001% methyl jasmonate. Leaves were harvested 1.5 hours and 6 hours after spraying

S10:

Oxidative stress: Seeds of *Arabidopsis thaliana* (wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were transferred to a growth chamber after three days. The intensity of light in the growth chamber was 7000-8000 LUX, temperature was 22°C, with 16 h light and 8 h dark. After 24 days, a few leaves were inoculated with a mixture of 2.5 mM D-glucose, 2.5 U/mL glucose oxidase in 20 mM sodium phosphate buffer pH 6.5. After an hour, 3 hours, or 5 hours after inoculation, whole plant, except for the inoculated leaves, was harvested. This sample was mixed with sample from plants that were sitting in full sun (152,000 LUX) for 2 hours or four hours.

Drought stress: Seeds of *Arabidopsis thaliana* (wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were transferred to a growth chamber after three days. The intensity of light in the growth chamber was 7000-8000 LUX, temperature was 22°C, with 16 h light and 8 h dark. After 20 days, aerial tissues were harvested and left to dry in 3MM Whatman paper for 1 hour or 4 hours.

Oxygen stress: Seeds of *Arabidopsis thaliana* (wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were transferred to a growth chamber after three days. The intensity of light in the growth chamber was 7000-8000 LUX, temperature was 22°C,

with 16 h light and 8 h dark. After 21 days, the plant was flooded by immersing its pot in a beaker of tap water. After 6 days, the upper tissues were harvested.

<u>S11: Heat-treated light grown seedlings:</u> Seeds of *Arabidopsis thaliana* (wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were transferred to a growth chamber after three days. The intensity of light in the growth chamber was 7000-8000 LUX, temperature was 22°C, with 16 h light and 8 h dark. Over a 5 hour period, the temperature was raised to 42°C at the rate of approximately 4°C per hour. After 1 hour at 42°C, the aerial tissues were collected. This sample was mixed with an equal volume of sample that went through a heat-recovery treatment namely bringing down the temperature to 22°C from 42°C over a 5 hour period at the rate of 4°C per hour.

<u>Cold-treated light grown seedlings:</u> Seeds of *Arabidopsis thaliana* (wassilewskija) were sown in pots and left at 4°C for two to three days to vernalize. They were transferred to a growth chamber after three days. The intensity of light in the growth chamber was 7000-8000 LUX, temperature was 22°C, with 16 h light and 8 h dark. After 18 days, the plant was transferred to 4°C for an hour before the aerial tissues were harvested. This sample was mixed with aerial tissues from another plant that was transferred to 4°C for 27 hours before being harvested.

<u>Analysis of Data:</u>

<u>Intensity</u>: The intensity of the band corresponds to the value in each lane marked S1, S2 etc.

<u>P-values</u>: The data shows P- values of each of the samples 1-11. P-values are calculated using the following formula 2*(1-NORMDIST(ABS(Sx-AVERAGE(of S1 to S11, not including Sx))/STDEV(of S1 to S11 not including Sx),0,1,TRUE)) using Excel functions.

The equivalent mathematical formula of P-value is as follows:

$\int \varphi(x)\,dx$, integrated from a to ∞,

where $\varphi(x)$ is a normal distribution:

where $a = \underline{|Sx-\mu|}$

σ(S1...S11, not including Sx);

where μ = is the average of the intensities of all samples except Sx,

$$= \frac{(\Sigma S1...Sn)-Sx}{n-1}$$

where σ(S1...S11, not including Sx) = the standard deviation of all sample intensities except Sx.

Results:

The results are shown in the MA_diff and/or AFLP_diff AFLP_int and AFLP_diff tables.

EXAMPLE 5: TRANSFORMATION OF CARROT CELLS

Transformation of plant cells can be accomplished by a number of methods, as described above. Similarly, a number of plant genera can be regenerated from tissue culture following transformation. Transformation and regeneration of carrot cells as described herein is illustrative.

Single cell suspension cultures of carrot (*Daucus carota*) cells are established from hypocotyls of cultivar Early Nantes in B_5 growth medium (O.L. Gamborg et al., *Plant Physiol.* 45:372 (1970)) plus 2,4-D and 15 mM $CaCl_2$ (B_5 -44 medium) by methods known in the art. The suspension cultures are subcultured by adding 10 ml of the suspension culture to 40 ml of B_5-44 medium in 250 ml flasks every 7 days and are maintained in a shaker at 150 rpm at 27 °C in the dark.

The suspension culture cells are transformed with exogenous DNA as described by Z. Chen et al. *Plant Mol. Bio.* 36:163 (1998). Briefly, 4-days post-subculture cells are incubated with cell wall digestion solution containing 0.4 M sorbitol, 2% driselase, 5mM MES (2-[N-Morpholino] ethanesulfonic acid) pH 5.0 for 5 hours. The digested cells are pelleted gently at 60 xg for 5 min. and washed twice in W5 solution containing 154 mM NaCl, 5 mM KCl, 125 mM $CaCl_2$ and 5mM glucose, pH 6.0. The protoplasts are suspended in MC solution containing 5 mM MES, 20 mM $CaCl_2$, 0.5 M mannitol, pH 5.7 and the protoplast density is adjusted to about 4×10^6 protoplasts per ml.

15-60 μg of plasmid DNA is mixed with 0.9 ml of protoplasts. The resulting suspension is mixed with 40% polyethylene glycol (MW 8000, PEG 8000), by gentle inversion a few times

at room temperature for 5 to 25 min. Protoplast culture medium known in the art is added into the PEG-DNA-protoplast mixture. Protoplasts are incubated in the culture medium for 24 hour to 5 days and cell extracts can be used for assay of transient expression of the introduced gene. Alternatively, transformed cells can be used to produce transgenic callus, which in turn can be used to produce transgenic plants, by methods known in the art. See, for example, Nomura and Komamine, *Plt. Phys.* 79:988-991 (1985), *Identification and Isolation of Single Cells that Produce Somatic Embryos in Carrot Suspension Cultures.*

EXAMPLE 6: PHENOTYPE SCREENS AND RESULTS

A: Triparental Mating and Vacuum Infiltration Transformation of Plants

Standard laboratory techniques are as described in Sambrook et al. (1989) unless otherwise stated. Single colonies of *Agrobacterium* C58C1Rif, *E. coli* helper strain HB101 and the *E. coli* strain containing the transformation construct to be mobilized into *Agrobacterium* were separately inoculated into appropriate growth media and stationary cultures produced. 100 μl of each of the three cultures were mixed gently, plated on YEB (5g Gibco beef extract, 1g Bacto yeast extract, 1g Bacto peptone, 5g sucrose, pH 7.4) solid growth media and incubated overnight at 28°C. The bacteria from the triparental mating were collected in 2 ml of lambda buffer (20 mM Tris (pH 7.5), 100 mM NaCl, 10 mM $MgCl_2$) and serial dilutions made. An aliquot of the each dilution was then plated and incubated for 2 days at 28°C on YEB plates supplemented with 100 μg/ml rifampicin and 100 μg/ml carbenicillin for calculation of the number of acceptor cells and on YEB plates supplemented with 100 μg/ml rifampicin, 100 μg/ml carbenicillin and 100 μg/ml spectinomycin for selection of transconjugant cells. The cointegrate structure of purified transconjugants was verified via Southern blot hybridization.

A transconjugant culture was prepared for vacuum infiltration by inoculating 1 ml of a stationary culture arising from a single colony into liquid YEB media and incubating at 28°C for approximately 20 hours with shaking (220 rpm) until the OD taken at 600 nm was 0.8-1.0. The culture was then pelleted (8000 rpm, 10 min, 4°C in a Sorvall SLA 3000 rotor) and the bacteria resuspended in infiltration medium (0.5X MS salts, 5% w/v sucrose, 10 μg/l BAP, 200 μl/l Silwet L-77, pH 5.8) to a final OD_{600} of 1.0. This prepared transconjugant culture was used within 20 minutes of preparation.

Wild-type plants for vacuum infiltration were grown in 4-inch pots containing Metromix 200 and Osmocote. Briefly, seeds of *Arabidopsis thaliana* (ecotype Wassilewskija) were sown in pots and left at 4°C for two to four days to vernalize. They were then transferred to 22-25°C and grown under long-day (16 hr light: 8 hr dark) conditions, sub-irrigated with water. After bolting, the primary inflorescence was removed and, after four to eight days, the pots containing the plants were inverted in the vacuum chamber to submerge all of the plants in the prepared transconjugant culture. Vacuum was drawn for two minutes before pots were removed, covered with plastic wrap and incubated in a cool room under darkness or very low light for one to two days. The plastic wrap was then removed, the plants returned to their previous growing conditions and subsequently produced (T1) seed collected.

B: Selection of T-DNA Insertion Lines

Approximately 10,750 seeds from the initial vacuum infiltrated plants were sown per flat of Metromix 350 soil. Flats were vernalized for four to five days at 4°C before being transferred to

22-25°C and grown under long-day (16 hr light: 8 hr dark) conditions, sub-irrigated with water. Approximately seven to ten days after germination, the (T1) seedlings were sprayed with 0.02% Finale herbicide (AgrEvo). After another five to seven days, herbicide treatment was repeated. Herbicide resistant T1 plants were allowed to self-pollinate and T2 seed were collected from each individual. In the few cases where the T1 plant produced few seed, the T2 seed was planted in bulk, the T2 plants allowed to self-pollinate and T3 seed collected.

C: Phenotype Screening

Approximately 40 seed from each T2 (or T3) line were planted in a 4-inch pot containing either Sunshine mix or Metromix 350 soil. Pots were vernalized for four to five days at 4°C before being transferred to 22-25°C and grown under long-day (16 hr light: 8 hr dark) conditions, sub-irrigated with water. A first phenotype screen was conducted by visually inspecting the seedlings five to seven days after germination and aberrant phenotypes noted. Plants were then sprayed with Finale herbicide within four days (i.e. about seven to nine days after germination).

The second visual screen was conducted on surviving T2 (or T3) plants about sixteen to seventeen days after germination and the final screen was conducted after the plants had bolted and formed siliques. Here, the third and fourth green siliques were collected and aberrant phenotypes noted. The Knock-in and Knock-out Tables contain descriptions of identified phenotypes.

Alternative, seed were surface sterilized and transferred to agar solidified medium containing Murashige and Skoog salts (1X), 1% sucrose (wt/v) pH 5.7 before autoclaving. Seed were cold treated for 48 hours and transferred to long days [16 hours light and 8 hours dark], 25°C. Plants were screened at 5 and 10 days.

In another screen, seed were surface sterilized and transferred to agar solidified medium containing Murashige and Skoog salts (1X), and combinations of various nitrogen and sucrose amounts as specified below::

Medium 1: no sucrose, 20.6 mM NH_4NO_3, 18.8 mM KNO_3;

Medium 2: 0.5% sucrose, 20.6 mM NH_4NO3, 18.8 mM KNO_3;

Medium 3: 3% sucrose, 20.6 mM NH_4NO_3, 18.8 mM KNO_3;

Medium 4: no sucrose, 20.6 μM NH_4NO_3, 18.8 μM KNO_3;

Medium 5: 0.5% sucrose, 20.6 μM NH_4NO_3, 18.8 μM KNO_3; and

Medium 6: 3% sucrose, 20.6 μM NH_4NO_3, 18.8 μM KNO_3.

The 0.5% sucrose was the control concentration for the sucrose. The low nitrogene, 20.6 μM NH_4NO_3, 18.8 μM KNO_3, is the control for the nitrogen. Seed were cold treated for 48 hours and transferred to long days [16 hours light and 8 hours dark], 25°C. Plants were screened at 2, 5, and 10 days.

D: TAIL-PCR and Fragment Sequencing

Rosette leaves were collected from each putative mutant and crushed between parafilm and FTA paper (Life Technologies). Two $2mm^2$ hole punches were isolated from each FTA sample and washed according to the manufacturer's instructions by vortexing with 200 ul of the provided FTA purification reagent. The FTA reagent was removed and the washing procedure repeated two more times. The sample was then washed twice with 200 ul of FTA TE (10 mM Tris, 0.1 mM EDTA, pH 8.0) and vortexing prior to PCR.

Primers used for TAIL-PCR are as follows:

AD2: 5' NGTCGASWGANAWGAA 3' (128-fold degeneracy)

S = G or C, W = A or T, and N = A, G, C, or T

LB1: 5' GTTTAACTGCGGCTCAACTGTCT 3'

LB2: 5' CCCATAGACCCTTACCGCTTTAGTT 3'

LB3: 5' GAAAGAAAAAGAGGTATAACTGGTA 3'

The extent to which the left and right borders of the T-DNA insert were intact was measured for each line by PCR. The following components were mixed for PCR: 1 $2mm^2$ FTA sample, 38.75 µl distilled water, 5 µl 10X Platinum PCR buffer (Life Technologies), 2 µl 50 mM $MgCl_2$, 1 µl 10 mM dNTPs, 1 µl 10 µM primer LB1 (or RB1 for analysis of the right border), 1 µl 10 µM primer LB3R (or RB3R for analysis of the right border) and 1.25 U Platinum Taq (Life Technologies). Cycling conditions were: 94°C, 10 sec.; thirty cycles of 94°C, 1 sec. - 54°C, 1 sec. - 72°C, 1 sec.; 72°C, 4 sec. The expected band size for an intact left border is bp, while an intact right border generates a bp band.

Fragments containing left or right border T-DNA sequence and adjacent genomic DNA sequence were obtained via PCR. First product PCR reactions use the following reaction mixture: 1 $2mm^2$ FTA sample, 12.44 µl distilled water, 2 µl 10X Platinum PCR buffer (Life Technologies), 0.6 µl 50 mM $MgCl_2$, 0.4 µl 10 mM dNTPs, 0.4 µl 10 µM primer LB1 (or RB1 for analysis of the right border), 3 µl 20 µM primer AD2 and 0.8 U Platinum Taq (Life Technologies). Cycling conditions for these reactions were: 93°C, 1 min.; 95°C, 1 min.; three cycles of 94°C, 45 sec. - 62°C, 1 min. - 72°C, 2.5 min.; 94°C,45 sec.; 25°C, 3 min.; ramp to 72°C in 3 min.; 72°C, 2.5 min.; fourteen cycles of 94°C, 20 sec. - 68°C, 1 min. - 72°C, 2.5 min. - 94°C, 20 sec.; - 68°C, 1 min. - 72°C, 2.5 min. - 94°C, 20 sec. - 44°C, 1 min. - 72°C, 2.5 min.; 72°C, 5 min.; end; ~4.5 hrs. For second product PCR reactions 1 µl of a 1:50 dilution of the first PCR product reaction was mixed with 13.44 µl distilled water, 2 µl 10X Platinum PCR buffer (Life

Technologies), 0.6 μl 50 mM $MgCl_2$, 0.4 μl 10 mM dNTPs, 0.4 μl 10 μM primer LB2 (or RB2 for analysis of the right border), 2 μl 20 μM primer AD2 and 0.8 U Platinum Taq (Life Technologies). Second product cycling conditions were: eleven cycles of 94°C, 20 sec. - 64°C, 1 min. - 72°C, 2.5 min. - 94°C, 20 sec. - 64°C, 1 min. - 72°C, 2.5 min. - 94°C, 20 sec. - 44°C, 1 min.; 72°C, 5 min.; end; ~3 hrs. Third product PCR reactions were prepared by first diluting 2 μl of the second PCR product with 98 μl of distilled water and then adding 1 μl of the dilution to 13.44 μl distilled water, 2 μl 10X Platinum PCR buffer (Life Technologies), 0.6 μl 50 mM $MgCl_2$, 0.4 μl 10 mM dNTPs, 0.4 μl 10 μM primer LB3 (or RB3 for analysis of the right border), 2 μl 20 μM primer AD2 and 0.8 U Platinum Taq (Life Technologies). Third product cycling conditions were: twenty cycles of 94°C, 38 sec. - 44°C, 1 min. - 72°C, 2.5 min.; 72°C, 5 min.; end; ~2 hrs. Aliquots of the first, second and third PCR products were electrophoresed on 1% TAE (40 mM Tris-acetate, 1 mM EDTA) to determine their size.

Reactions were purified prior to sequencing by conducting a final PCR reaction. Here, 0.25 μl Platinum PCR Buffer (Life Technologies), 0.1 μl 50 mM $MgCl_2$, 3.3 U SAP shrimp alkaline phosphatase, 0.33 U Exonuclease and 1.781 μl distilled water were added to a 5 μl third product and the reaction cycled at 37°C, 30 min.; 80°C, 10 min.; 4°C indefinitely.

Di-deoxy "Big Dye" sequencing was conducted on Perkin-Elmer 3700 or 377 machines.

KNOCK-IN EXPERIMENTS

For the following examples, a two-component system was constructed in a plant to ectopically express the desired cDNA.

First, a plant was generated by inserting a sequence encoding a transcriptional activator downstream of a desired promoter, thereby creating a first component where the desired promoter facilitates expression of the activator generated a plant. The first component also is referred to as the activator line.

Next, the second component is constructed by linking a desired cDNA to a sequence that the transcriptional activator can bind to and facilitate expression of the desired cDNA. The second component can be inserted into the activator line by transformation. Alternatively, the

second component can be inserted into a separate plant, also referred to as the target line. Then, the target and activator lines can be crossed to generate progeny that have both components.

Two component lines were generated by both means.

Part I - From crosses

Target lines containing cDNA constructs are generated using the Agrobacterium-mediated transformation. Selected target lines are genetically crossed to activation lines (or promoter lines). Generally, the promoter lines used are as described above. Evaluation of phenotypes is done on the resulting F1 progenies.

Part II - From Type I Supertransformation

Promoter activation lines (generally Vascular/Ovule/Young Seed/Embryo line, Seed/Epidermis/Ovary/Fruit line, Roots/Shoots/Ovule line, and Vasculature/Meristem are transformed with cDNA constructs using the Agrobacterium mediated transformation. Selected transformants (and their progenies) are evaluated for changes in phenotypes. The table for the knock-in of the Type I supertransformation comprises the following information

- Clone ID,
- Pfam,
- Gemini ID
- Trans. Unique ID (which indicates what promoter activation line was transformed
- S Ratio: segregation ratio after the transformed plants are selected for the marker.
- Assay
- Stage: phenotype was observed
- Feature: Where the phenotype was observed
- Phenotype

- P Ratio: phenotype ratio
- Comments

Part III - From Type II Supertransformation

Target lines generated using the procedure mentioned in Part I are transformed with T-DNA construct containing constitutive promoter. Selected transformants (and their progenies) are evaluated for changes in phenotypes.

An additional deposit of an *E. coli* Library, *E. coli*LibA021800, was made at the American Type Culture Collection in Manassas, Virginia, USA on February 22, 2000 to meet the requirements of Budapest Treaty for the international recognition of the deposit of microorganisms. This deposit was assigned ATCC accession no. PTA-1411. Additionaly, ATCC Library deposits; PTA-1161, PTA-1411 and PTA-2007 were made at the American Type Culture Collection in Manassas, Virginia, USA on; January 7, 2000, February 23, 2000 and June 8, 2000 respectively, to meet the requirements of Budapest Treaty for the international recognition of the deposit of microorganisms.

The invention being thus described, it will be apparent to one of ordinary skill in the art that various modifications of the materials and methods for practicing the invention can be made. Such modifications are to be considered within the scope of the invention as defined by the following claims.

Each of the references from the patent and periodical literature cited herein is hereby expressly incorporated in its entirety by such citation.